9-1-2018

Molecular Epidemiology, Ecology, and Evolution of Group A Streptococci

Debra E. Bessen
New York Medical College

Pierre R. Smeesters

Bernard W. Beall

Recommended Citation

Bessen, D. E., Smeesters, P., & Beall, B. (2018). Molecular Epidemiology, Ecology, and Evolution of Group A Streptococci. *Microbiology Spectrum, 6* (5). https://doi.org/10.1128/microbiolspec.CPP3-0009-2018

Microbiology Spectrum
American Society for Microbiology Press

Molecular Epidemiology, Ecology, and Evolution of Group A Streptococci

DEBRA E. BESSEN,[1] PIERRE R. SMEESTERS,[2] and BERNARD W. BEALL[3]

[1]Department of Microbiology and Immunology, New York Medical College, Valhalla, NY 10595; [2]Department of Pediatrics, Queen Fabiola Children's University Hospital, and Molecular Bacteriology Laboratory, Université Libre de Bruxelles, Brussels, 1020, Belgium; [3]National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333

ABSTRACT The clinico-epidemiological features of diseases caused by group A streptococci (GAS) is presented through the lens of the ecology, population genetics, and evolution of the organism. The serological targets of three typing schemes (M, T, SOF) are themselves GAS cell surface proteins that have a myriad of virulence functions and a diverse array of structural forms. Horizontal gene transfer expands the GAS antigenic cell surface repertoire by generating numerous combinations of M, T, and SOF antigens. However, horizontal gene transfer of the serotype determinant genes is not unconstrained, and therein lies a genetic organization that may signify adaptations to a narrow ecological niche, such as the primary tissue reservoirs of the human host. Adaptations may be further shaped by selection pressures such as herd immunity. Understanding the molecular evolution of GAS on multiple levels—short, intermediate, and long term—sheds insight on mechanisms of host-pathogen interactions, the emergence and spread of new clones, rational vaccine design, and public health interventions.

ECOLOGY: HOST TISSUES AND DISEASES

The group A streptococcus (GAS; *Streptococcus pyogenes*) is a free-living organism. Its ecological niche appears to be quite narrow, and its only known biological host of import is humans. There may be occasional or very rare natural infections in non-human primates and other mammals (1, 2).

The primary sites for GAS colonization of the human host involve two tissues: the oropharyngeal mucosal epithelium of the upper respiratory tract (URT) and the superficial layers of the epidermis. GAS have no known environmental reservoir of significance, and transmission is almost exclusively person to person, via either respiratory droplets or direct contact. Thus, the throat and skin of the human host are the primary habitats for GAS. It is at these two tissue sites that GAS undergo successful reproductive growth and transmission of progeny to new hosts (Fig. 1).

Streptococcal pharyngitis and impetigo are superficial, self-limiting infections that usually cause a mild illness that resolves within 2 weeks, even if left untreated. Both are primarily diseases of childhood. Once colonization is established at the URT, the organism can cause a symptomatic pharyngitis, with or without tonsillitis. Symptomatic infection caused by GAS is typically characterized by a copious purulent exudate. Oropharyngeal

Received: 1 April 2018, **Accepted:** 13 April 2018,
Published: 6 September 2018

Editors: Vincent A. Fischetti, The Rockefeller University, New York, NY; Richard P. Novick, Skirball Institute for Molecular Medicine, NYU Medical Center, New York, NY; Joseph J. Ferretti, Department of Microbiology & Immunology, University of Oklahoma Health Science Center, Oklahoma City, OK; Daniel A. Portnoy, Department of Molecular and Cellular Microbiology, University of California, Berkeley, Berkeley, CA; Miriam Braunstein, Department of Microbiology and Immunology, University of North Carolina-Chapel Hill, Chapel Hill, NC, and Julian I. Rood, Infection and Immunity Program, Monash Biomedicine Discovery Institute, Monash University, Melbourne, Australia

Citation: Bessen DE, Smeesters PR, Beall BW. 2018. Molecular epidemiology, ecology, and evolution of group A streptococci. *Microbiol Spectrum* 6(5):CPP3-0009-2018. doi:10.1128/microbiolspec.CPP3-0009-2018.

Correspondence: Debra E. Bessen, debra_bessen@nymc.edu

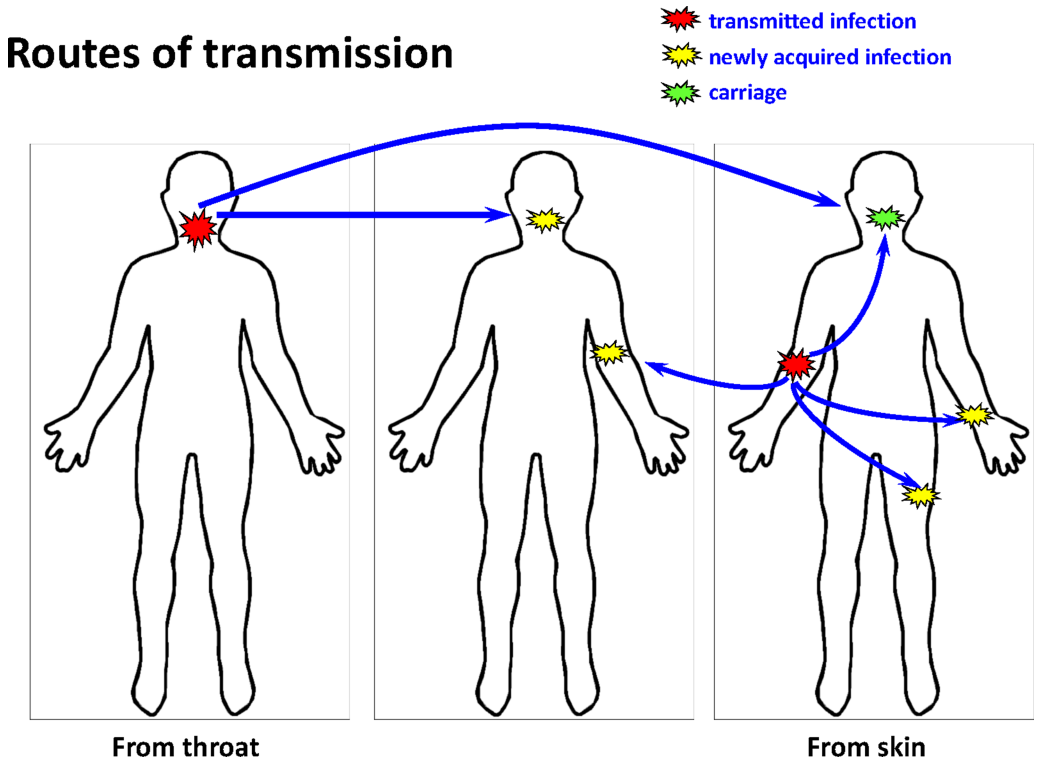


FIGURE 1 Transmission routes for GAS. From the throat, GAS can transmit via the respiratory route to a new host, where it either causes pharyngitis or persists in a quiescent carrier state. In a carrier state, the organism is presumed to be only weakly transmissible. Transmission from the throat to skin is relatively rare. From an impetiginous skin lesion, the organism can be transmitted by direct contact to the slightly damaged skin of a new host or to other damaged skin sites on the same host, causing multiple skin lesions. GAS can also be transmitted from a skin lesion to the throat of the same host, but it is widely assumed to enter a carrier state and is only weakly transmissible. From either the throat or skin, the organism can invade normally sterile deeper tissue, but this is rare relative to superficial infections.

infection can be cleared via a protective antibody that is produced by the host in response to infection. Alternatively, a "clinically inapparent" infection can arise wherein the infected individual lacks obvious clinical symptoms of illness, yet a specific immune response is mounted and directed to antigens of the infecting organism (3).

Another type of GAS-human interaction that can occur at the URT is asymptomatic carriage, where a vigorous host immune response to many/most GAS antigens is lacking (4). The carrier state can persist for weeks or months, and the organism (presumably) remains capable of transmission to a new host. However, relative to acute pharyngitis, the number of organisms present during carriage may be vastly reduced, and GAS are often in an altered dormant-like state signified by a relative lack of M protein expressed on the cell surface. Intracellular invasion of epithelial cells by GAS may contribute to persistent infection (5). Microcolony and biofilm formation, and the bacterial communities that arise, may also be a key part of the ecology of the GAS carrier state (6).

At the skin, GAS survive and replicate below the dry cornified layer, provoking a strong inflammatory response that gives rise to the purulent lesions of nonbullous impetigo, a form of pyoderma. Recovery of GAS from normal skin can precede impetigo by ~10 days (7, 8). However, it is not entirely clear whether there exists stable GAS colonization and/or a true long-term carrier state at the skin; uncertainty stems from the duration of GAS association with normal human skin in the absence of inflammation (i.e., disease), which may be transient (9). Breaks in the stratum corneum, which can be quite minor and barely noticeable, are a requirement for GAS to establish a firm foothold at this tissue site and cause disease. The scabies mite infests the epidermal layer just below the stratum corneum; scabies can cause intense

itching and is a key risk factor in the development of GAS impetigo in many tropical communities (10–12). Also, it is not uncommon for *Staphylococcus aureus* to be a copathogen along with GAS in impetiginous skin lesions (13, 14). At least some strains of GAS that cause impetigo can subsequently spread from skin lesions to colonize the URT, peaking at 2 to 3 weeks post-skin infection (7, 8) (Fig. 1).

The relative incidence of GAS disease varies throughout the world in accordance with both season and locale. In the temperate regions of North America and Europe, pharyngitis is highly prevalent during the cold winter months, whereas impetigo thrives during warm, humid weather and is far less prevalent. In many tropical regions, such as the Northern Territory of Australia, GAS impetigo is far more common than pharyngeal infection among the indigenous population, and there are no discrete seasonal peaks in the incidence of disease (10). Climate and the level of hygiene strongly influence the prevalence of impetigo, whereas indoor crowding is a major risk factor for URT infection. These risk factors largely reflect the primary modes of transmission.

Worldwide there are an estimated >616 million cases of GAS pharyngitis per year (15), and >162 million cases of impetigo at any given time (16). In a study conducted ~40 years ago in the United States, for children between the ages of 5 and 7 years, it was estimated that an average of half experience one GAS infection annually (17). During the peak season, asymptomatic carriage rates at the URT range from (very) roughly 15 to 20% or higher; this value depends on the exact cohort (18). Bacteria tend to attain high numbers and densities during purulent infection. Large bacterial counts yield large infective doses, which in turn can lead to the exposure of several susceptible hosts with the minimum infective dose, thereby augmenting transmission and spread.

A primary reason why GAS is a major public health problem is because of its ability to cause invasive (iGAS) disease and trigger nonsuppurative (i.e., nonpus-forming) sequelae. The latter diseases are of an autoimmune (acute rheumatic fever [ARF]) or immune complex (acute glomerulonephritis) nature. Collectively, the severe GAS diseases are responsible for >500,000 deaths each year (15).

In iGAS disease, bacteria gain access to deep tissue that is normally sterile. Consistent with pharyngitis providing the major GAS reservoir in temperate regions, iGAS disease is also more prevalent in these regions during the winter months (19). Intracellular invasion or paracellular transport of GAS at the URT may facilitate its access to the bloodstream. iGAS infection can be particularly severe in cases of necrotizing fasciitis and toxic shock syndrome. Although iGAS disease is associated with elevated rates of morbidity and mortality, its frequency of occurrence is relatively low. Population-based surveys of iGAS disease over the past 2 decades estimate the annual incidence at ~3 to 4 per 100,000 people per year in the United States and Europe, although more recently, the incidence of iGAS disease in the United States has increased to ~5 per 100,000 people per year (20); worldwide, there are an estimated 663,000 new iGAS cases per year (15). Thus, the overall incidence of iGAS disease is ~1,000-fold lower than the incidence of new pharyngitis and impetigo infections. Invasion of normally sterile tissue is often a result of "bacteremia without focus." When a lack of evidence for a superficial infection is coupled to GAS growth being largely restricted to deep tissue, and the patient is sequestered and nonambulatory, iGAS disease becomes a transmission dead end for the infecting bacterium.

From the ecological perspective of long-term evolution, GAS fitness is largely shaped by adaptations to its primary tissue reservoirs; acquisition of traits that enhance fitness in iGAS disease is probably coincidental. In North America, the composition of the iGAS strain population largely overlaps with the population of GAS strains recovered from superficial (pharyngitis) infections (21), although a few strains appear to have invasive indices at the more extreme high or low ends (22).

Cutaneous wound infections can be considered a mild form of iGAS disease, wherein the histopathology is distinct from nonbullous impetigo. Depending on the study, wound infections are occasionally classified under the broader umbrellas of "pyoderma" or "skin lesions." Ecthyma is a severe form of impetigo with extensive dermis involvement. Impetigo largely afflicts younger children and is mostly prevalent in communities with a warm humid climate (tropics, summer), whereas in contrast, GAS wound infections occur in all age groups and in all geographic regions, including those where URT disease predominates.

Transmission of GAS leading to cutaneous wound infections can occur from one tissue to another site within an individual (e.g., URT → broken skin), through shared skin-penetrating devices (e.g., needles), and possibly by direct person-to-person contact via open wounds. Skin contact may be an important mode of transmission in a recent GAS outbreak among adults that include intravenous drug users (IVDUs) and homeless people in a temperate zone (23, 24).

ECOLOGY: TRANSMISSION

The superficial nature of pharyngitis and impetigo provides GAS with an easy exit from infected hosts and an easy entrance into new hosts. Transmission is strongly influenced by the GAS agent and human host, coupled with both microenvironmental and macroenvironmental factors. The dynamics of both the GAS and human host populations largely dictate the scope of GAS disease outbreaks and epidemics.

Macroenvironment Risk Factors

As discussed above ("Ecology: Host Tissues and Diseases"), the primary environmental risk factors for GAS pharyngitis and impetigo are cold winters and tropical conditions, respectively. Copathogens that may elevate the risk for GAS infection include the scabies mite and *S. aureus* for impetigo and, possibly, undefined infectious agents for pharyngitis.

GAS-Microenvironment Factors

The microenvironments at the throat and skin differ from one another in numerous ways. Also, each tissue undergoes fluctuations in local conditions as a GAS infection evolves. To initiate infection, the GAS organism needs to be reasonably well adapted to the local tissue microenvironment. Distinctions between classic throat and skin strains of GAS (see "Epidemiology: Global Trends" below) are probably rooted, at least in part, in host tissue-specific adaptations. GAS cell surface components implicated in transmission and/or mediating colonization in a new host include the hyaluronic acid capsule and pili, among numerous other extracellular proteins (25–29). Yet the recent emergence and successful spread of *emm89* iGAS strains is coupled to a loss in capsule (30, 31) (see "Epidemiology: Building Evolutionary Models" below). Conceivably, GAS transmission factors are strain specific.

At the URT, GAS initially encounter saliva and mucus secretions that contain mucins, secretory IgA, digestive enzymes, and antimicrobial peptides. At the broken surface just below the cornified layer of the skin, GAS face the early stages of wound healing: an inflammatory response in the epidermis is characterized by infiltration of low numbers of polymorphonuclear leukocytes and small amounts of plasma containing Igs and complement, plus antimicrobial peptides. As the infection evolves, the bacteria deploy virulence factors that are able to thwart host defenses, allowing GAS to undergo a net increase in growth, with a concomitant increase in infiltration by inflammatory cells leading to the accumulation of a purulent exudate (32).

Competition with normal flora is a key determinant of GAS colonization at the URT. In general terms, competition can be stiffest between highly related organisms. The closest genetic relative of GAS/*S. pyogenes* that is also a common human URT colonizer is *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE). SDSE usually have cell wall-associated group C or G group carbohydrate and are largely restricted to humans, with >60% genome sequence homology to *S. pyogenes* (33, 34). A few uncommon lineages of SDSE have group A (CHO-A) or group L carbohydrate (thus, the term "GAS" is largely equivalent to "*S. pyogenes*," and in this report, the two terms are completely equivalent unless otherwise noted). Recently, a widely disseminated and longstanding iGAS lineage of SDSE was described that appears to have arisen from a recombination event involving CHO-A biosynthesis genes from *S. pyogenes* (35).

SDSE bacteria are largely commensals of the oropharynx but are capable of causing many of the diseases attributed to *S. pyogenes* (36, 37). Like its close cousin, SDSE expresses capsules composed of hyaluronic acid, cell surface proteins that include M protein and T-antigens, and exotoxins such as streptolysin O; however, SDSE lacks some critical *S. pyogenes* virulence factors, such as the secreted cysteine proteinase SpeB. Besides the potential sharing of human host receptors for adherence to the epithelium, relatively little is known about how these two species—GAS and SDSE—compete for nutrients and other critical resources (38). Signaling between GAS, SDSE, and other streptococcal species has been demonstrated (39), and GAS are susceptible to killing by bacteriocins produced by other streptococcal species (40). Much remains to be uncovered regarding the role of the microbiome in shaping the success of GAS in colonizing and/or causing infection at the URT.

GAS physiology is heavily influenced by the nature and concentration of carbon sources (e.g., glucose), oxygen, and its toxic byproducts, among numerous other factors present in different tissues under shifting conditions (41, 42). Changes in the local microenvironment, coupled with increased bacterial density and nutrient depletion as the organisms reproduce and yield new progeny, impact the transcriptional program of the GAS cell, which in turn directs activation or repression of numerous genes associated with pathogenesis, and presumably transmissibility as well (43).

Acquisition of GAS at the URT does not necessarily lead to an infection, and in the asymptomatic carrier state, the local microenvironment is substantially different from the purulent environs. In the carrier state,

virulence factor gene expression and the requirements for adaptation appear to be quite different (e.g., 44, 45).

Human (Host) Factors

In general terms, successful transmission of a pathogen depends on the availability of susceptible hosts, as well as the ratio of susceptible to resistant hosts in the context of the number of person contacts and potential exposures (46). Host genetic susceptibility (or resistance) to superficial GAS infections has not been characterized, and evidence for its existence is largely lacking.

Highly efficacious immunological protection against GAS infection is strain specific (i.e., M type-specific) (47), although a host immune response directed toward numerous other antigenic targets can confer at least partial protection (48, 49). Herd immunity is probably a critical host determinant for circumscribing the degree to which a given GAS strain can successfully transmit and spread within a human population. Thus, the possibility of future immunization with an M type-based vaccine is likely to reshape the composition of the GAS strain population (see "Epidemiology: Strategies for M Protein-Based Vaccines" below).

GAS (Agent) Factors

A key open question is whether GAS engage in a form of evolutionary bet-hedging via transmitting a bolus of organisms with a diverse array of transcriptional profiles, to increase the odds that at least one physiological form will colonize and thrive within an individual host. Alternatively, perhaps there exists a singular transcriptional profile that is deemed optimal for transmission within the wider human host population. Or perhaps a mixture of both strategies is used. Mutations in transcriptional regulatory genes and/or promoter regions may enhance the transmissibility of GAS (see "Evolution: Short Term" below).

GAS exhibit very high diversity in their cell surface proteins, in terms of both the presence or absence of a given protein, and in the sequence heterogeneity among products of the same locus. Several of the heterologous cell surface proteins may impart tissue-specific adaptations (reviewed in reference 50, 51). They also serve as key epidemiological markers (see next section). Genetic changes that lead to alterations in the antigenic composition of the GAS cell surface may aid the organism in escaping herd immunity.

EPIDEMIOLOGY: MOLECULAR MARKERS

The use of serological typing to categorize the pathogens of the *Streptococcus* genus (i.e., cell wall group carbohydrate), as well as the strains within a pathogenic streptococcal species, has a long and deep history. Unlike pneumococci, GAS lack a structurally diverse and immunogenic polysaccharide capsule. Instead, GAS express several cell surface proteins which collectively impart a broad repertoire of antigenic diversity. Underlying the serological phenotype is an extensive history of horizontal gene transfer (HGT) and recombination among the cell surface protein genes, plus accumulated point mutations, that has been shaped (at least in part) by the strong diversifying selection pressures of the host immune response. The antigenic structure of GAS cell surface proteins, and the underlying genotype of both the cell surface protein and core housekeeping genes, is highly complex.

Numerous typing schemes have been used to aid in understanding the epidemiology and underlying organization of the *S. pyogenes* species. A summary of typing (and related) schemes, and their applications and limitations, is provided in Table 1. Each epidemiological molecular marker is discussed in detail below.

M Protein

During the 1920s, work began that aimed at understanding the basis for protective immunity to GAS infection. Antibodies raised to extractable surface antigens, called M proteins, led to opsonophagocytosis of the strain from which the M protein was derived (52). However, antibody directed to the M protein of one organism often failed to confer protection against many other GAS isolates tested. A serological typing scheme arose through the development of antibodies directed to M proteins of different isolates. More than 80 distinct M types were identified, and protective immunity to GAS infection is, in large part, M type specific.

Copurifying with the M type-specific material was M-associated protein (MAP), which has two major antigenic forms (53, 54). The MAP I antigen was largely associated with ARF isolates, and MAP II was highly prevalent among serum opacity factor (SOF)-producing organisms. The MAP antigens were subsequently identified as part of the conserved C repeat region (CRR) of M protein (55), which occupies the COOH-terminal half of the surface-exposed portion of M protein. The CRR has two major antigenic forms known as class I and class II (56). The structure of M protein is depicted in Fig. 2A.

emm types

A nucleotide sequence-based *emm* typing scheme that closely parallels the M serotype scheme was developed ~2 decades ago (57). *emm* typing is based on the extensive

TABLE 1 Epidemiological typing methods for group A streptococci[a]

Scheme	Description of method	Applications	Limitations
M serotyping	Immunoprecipitation, using hyperimmune rabbit antisera	No longer in wide use	
emm typing	Sequencing, yielding >92% nt identity over the first 90 bases encoding the processed (mature) M protein of an *emm* type reference strain	Most commonly used typing method; highly reproducible; low cost; large database at CDC; can be used to infer *emm* clusters and *emm* pattern groups with reasonable accuracy	Based on oligonucleotide primers (via PCR); needs better integration with whole-genome sequencing projects
emm subtyping	Sequencing of 180 nt corresponding to the 10 COOH-terminal codons of the signal sequence and first 50 codons of the mature M protein	Further stratifies *emm* types into unique alleles; performed in parallel with *emm* typing; large database and BLAST server are maintained at the CDC	
emm pattern grouping	PCR-based mapping, based on 3′ ends of *emm* and *emm*-like genes	Assignment of organisms of known *emm* types into groupings that correspond reasonably well to tissue site preferences for superficial infection at the throat and/or skin	Inferences based on *emm* type may occasionally be incorrect; ~30% of *emm* types await *emm* pattern assignment; PCR-based mapping is laborious and limited by design of "universal" oligonucleotides
emm cluster typing	Based on phylogeny and four bioinformatic criteria, related to surface-exposed portion of M proteins; corresponds to binding by several human serum proteins	Used to extend information on *emm* typing; can be easily inferred from *emm* type; binding motifs can be used to infer binding properties of M proteins	Inferences based on *emm* type may occasionally be incorrect; ~29% of *emm* types await *emm* cluster assignment
SOF typing	Inhibition of the serum opacity reaction by hyperimmune rabbit serum	No longer in wide use	
sof typing	Sequencing of ~700 bp region near 5′ end of *sof*	Data can be generated though sequencing of PCR products or whole-genome sequencing	*sof* type assignment needs to undergo standardization; evidence for extensive recombination within the *sof* type region
T serotype	Bacterial cell agglutination, using hyperimmune rabbit antisera	In limited use by specialized laboratories, to supplement *emm* typing	Inconsistent results; labor-intensive; hyperimmune serum is limited and difficult to standardize
FCT region typing	PCR-based mapping; whole genome sequence	Can be used as an initial step in further characterization of pilin genes	Assignment of the FCT region needs to undergo further definition and standardization
FCT gene typing	Sequenced-based		Under development for all pilin subunit and adhesin genes
MLST	Nucleotide sequence of seven (partial) housekeeping genes, from PCR products or whole-genome sequences	Used to define "clones" (i.e., STs) and clonal complexes (i.e., CCs); often coupled with *emm* type; alleles are assigned via simple web-based user queries and curation; database of alleles, STs, and isolates is maintained on PubMLST server (Oxford)	Provides only a small snapshot of the core genome; isolate database is skewed toward unique genotypes (i.e., not population-based)
Whole-genome sequencing	Complete or partial genome assembly	Phylogenetic relationships of closely related organisms via the core genome sequence; investigations of local outbreaks and widespread epidemics; routine surveillance; pipelines for analysis of typing and antibiotic resistance genes	Core genome definition lacks standardization for the diverse range of strains in the species; assembly problems in regions of extensive sequence diversity and/or recombination; bioinformatic analysis can be time-consuming

[a]nt, nucleotide.

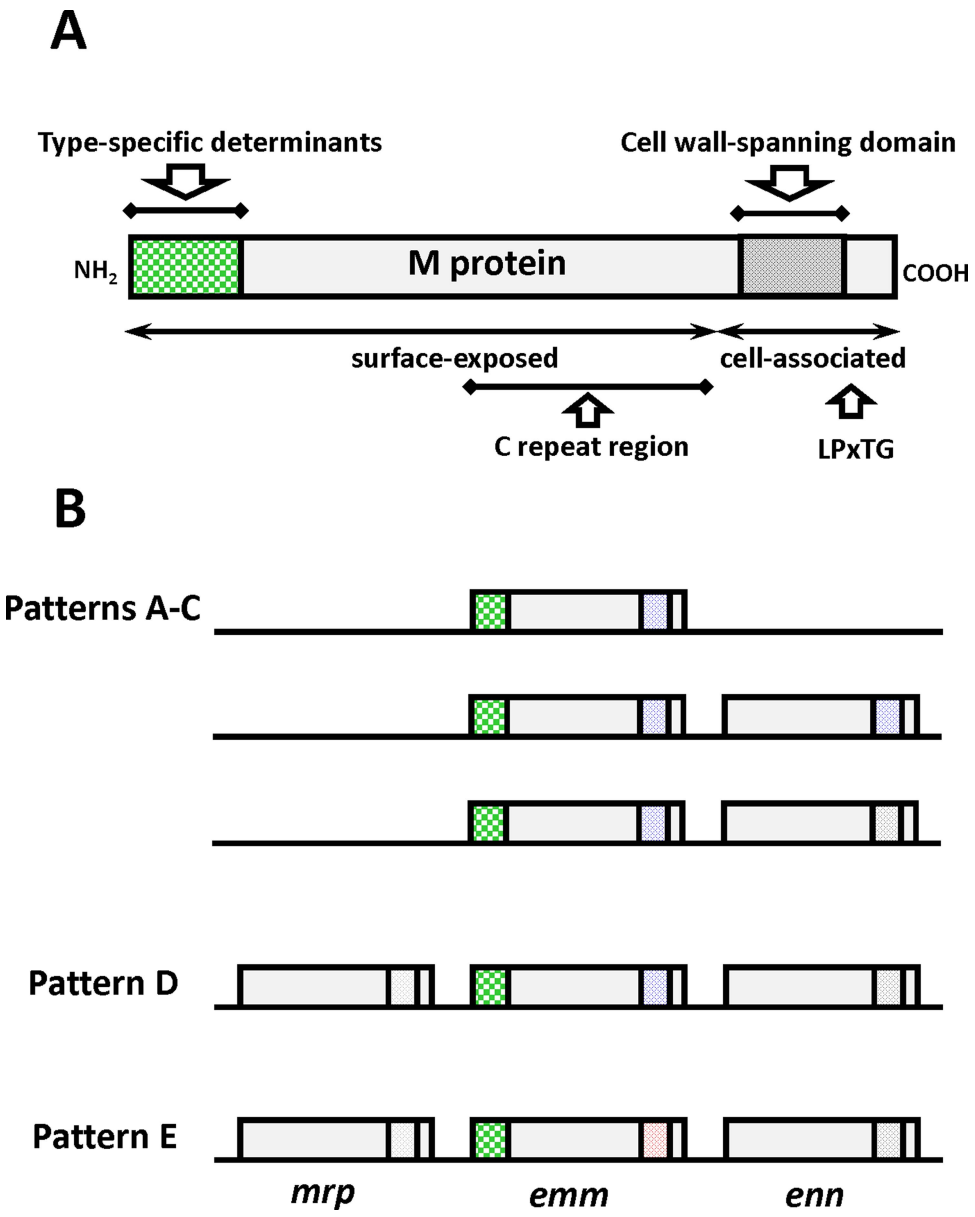


FIGURE 2 M protein structure and *emm* pattern gene arrangements. **(A)** Key features of M protein are shown, including the type-specific determinants, cell wall-spanning domain, and C repeat region. **(B)** Chromosomal arrangement of *emm* and the flanking *emm*-like genes (*mrp*, *enn*) gives rise to five *emm* patterns, which form three main groupings (A-C, D, E). Transcription of *emm* and *emm*-like genes is positively regulated by Mga, which is encoded by *mga*, which lies immediately upstream of the *emm* region.

nucleotide sequence differences at the 5′ end of the *emm* gene, whereby a unique *emm* type is defined as having <92% nucleotide identity to any other *emm* type, over the sequence corresponding to the first 30 codons of the mature M protein (Fig. 2A). Virtually all contemporary epidemiological studies define GAS isolates according to their *emm* type. Currently, there are ~243 recognized *emm* types (58).

Identification of the *emm* gene is not always obvious, and the problem can be due to paralogous *emm*-like genes mapping immediately upstream (*mrp*) or downstream (*enn*) of *emm* in numerous GAS strains (see next section; Fig. 2B). Confusion about *emm* type assignments has arisen wherein *enn* is directly fused to the 5′ end region of *emm* as the result of DNA excision arising from recombination between the *emm* and *enn* genes.

Among the ~243 *emm* types recognized to date, there are ~1,632 distinct allelic forms of the *emm* type-specific region, known as *emm* subtypes (58). The *emm* subtype alleles are based on sequence differences relative to the typing reference, within a 180-bp segment corresponding to the last 10 codons of the predicted signal peptide coding region and the first 50 codons corresponding to the NH_2-terminus of the mature M protein molecule (59).

emm patterns

Many strains of GAS have *emm*-like genes (*mrp*, *enn*) positioned immediately upstream and/or downstream of *emm* on the chromosome (60–63). Near the 3′ end of each *emm* and *emm*-like gene lies a region that encodes one of four divergent cell wall-spanning domains that can be distinguished by oligonucleotides (60, 64). Using a PCR-based mapping approach, the content of *emm* and *emm*-like genes and their arrangement on the chromosome was determined for GAS isolates of many different *emm* types. The resulting *emm* region maps are referred to as "*emm* patterns"; three major *emm* pattern groupings are recognized (Fig. 2B). M proteins of *emm* pattern E tend to be class II, whereas the *emm* patterns A-C and D groupings largely overlap with the class I CRR (60).

To date, the *emm* pattern genotype has been ascertained for GAS isolates of 170 of the ~243 recognized *emm* types (Table 2). With only two exceptions uncovered thus far (*emm54*, *emm218*), multiple isolates sharing the same *emm* type also share the same *emm* pattern grouping. The majority of *emm* types are represented by *emm* pattern groups D and E, accounting for >75% of the *emm* types analyzed.

emm clades and clusters

An *emm* cluster typing system was recently developed, based on the entire surface-exposed portion of M protein and its capacity to bind six human serum proteins (65). Phylogenetic analysis of >1,000 *emm* genes corresponding to 174 *emm* types reveals 2 main clades (X and Y), and 16 well-supported clusters, accounting for 82% of the *emm* types analyzed. The relationships between *emm* clades, clusters, and patterns for 174 *emm* types are illustrated in Table 3. Nearly all pattern E *emm* types (98%) belong to clade X, whereas 92% of pattern A-C *emm* types are clade Y.

In contrast to patterns A-C and E *emm* types, the pattern D *emm* types are present in both clades X and Y (Table 3) (65). The highly specialized plasminogen-binding *emm* cluster D4 of clade Y represents the largest *emm* pattern D grouping (n = 30 *emm* types). Within clade X, *emm* clusters E5 and E6 constitute a second group of pattern D *emm* types (n = 14); these two clusters include *emm* types of other pattern groups as well. Most of the remaining pattern D *emm* types are interspersed throughout clade Y (n = 16). Overall, the *emm* cluster scheme, based on the phylogeny of the entire surface-exposed portion of M protein, is highly consistent with the *emm* pattern scheme that is based on the COOH-terminal peptidoglycan-spanning domain of *emm* and *emm*-like flanking genes.

The *emm* cluster typing system is easy to implement since the *emm* cluster can be inferred from the *emm* type (66–69). Importantly, the *emm* type → *emm* cluster scheme is also highly informative of the functional capacity of the M protein molecule, based on consensus binding motifs for numerous human host plasma proteins (65).

TABLE 2 Distribution of *emm* types among *emm* pattern groups for 170 *emm* types.

emm pattern group	No. of *emm* types represented	% of total
A–C	35	20.6
D	64	37.6
E	67	39.4
Rearranged (REA)	2	1.2
A–C and D	2	1.2

M protein functional domains

As summarized in recent reviews (63, 70), the surface-exposed non-type-specific portion of M proteins (i.e., central region) contains functional domains that bind human host proteins such as IgG (various subclasses), IgA, plasminogen, and fibrinogen. The M type-specific determinants at the NH_2-terminus bind two complement regulators, C4b-binding protein (C4BP) and factor H-like 1.

Binding capacities strongly correlate with *emm* cluster (65). For example, binding of human IgG is mostly restricted to M proteins of *emm* clusters E1 through E6. Similarly, binding of C4BP is limited to M proteins of *emm* clusters E1, E3, E4, and E6. IgA binding is evident only in *emm* clusters E1, E4, and E6 products, whereas fibrinogen binding is restricted to M proteins of clade Y, and plasminogen binding is restricted *emm* cluster D4 (51, 65).

Although much less is known about the M-like proteins Mrp and Enn (Fig. 2B), binding properties shared with M protein have been described (reviewed in reference 63), and in some instances, they may use similar

TABLE 3 Distribution of *emm* clusters relative to *emm* pattern groups, for 174 *emm* types

emm cluster	*emm* clade	No. of *emm* types	No. of *emm* types: pattern A–C[a]	No. of *emm* types: pattern D	No. of *emm* types: pattern E	Other *emm* pattern (no. of *emm* types)[b]
E1	X	5	0	0	5	
E2	X	15	0	0	15	
E3	X	19	0	0	18	ND (1)
E4	X	17	0	0	17	
E5	X	7	2	3	0	REA (2)
E6	X	18	1	11	6	
singletons	X	4	0	1	3	
D1	Y	3	0	2	0	A–C and D (1)
D2	Y	5	0	5	0	
D3	Y	2	0	2	0	
D4	Y	31	0	30	0	ND (1)
D5	Y	3	1	2	0	
A-C1	Y	2	2	0	0	
A-C2	Y	2	2	0	0	
A-C3	Y	5	5	0	0	
A-C4	Y	5	5	0	0	
A-C5	Y	3	1	0	0	ND (2)
Singletons	Y	22	15	5	0	A–C and D (1); ND (1).
Outliers	Outlier	6	2	3	1	

[a]*emm* pattern assignments are based on references 65, 126, and 127.
[b]ND, not determined; REA, rearranged.

binding motifs. Additional investigation of the structure and functional binding properties of the M-like proteins, and structure of the *mrp* and *enn* genes, may eventually lead to a higher-resolution typing scheme that includes the entirety of the *emm* chromosomal region.

Serum Opacity Factor

SOF promotes the opacification of serum, wherein the serum opacity reaction can be neutralized by hyperimmune antiserum produced in animals (71). Not all GAS strains produce SOF, but SOF typing was often used as an adjunct for strains that posed challenges for M typing. Of note, SOF producers have a strong association with strains that harbor the MAP II antigen (72). In addition, SOF-producing strains tend to exhibit diminished beta-hemolysis (73).

More recent studies show SOF to be a proteinaceous virulence factor that is either bound to the GAS cell surface or secreted. SOF has an enzymatic domain that interacts with high-density lipoprotein, leading to the opacification of serum, plus a fibronectin-binding domain (74–77). The genes encoding M protein (*emm*) and SOF (*sof*) map ~15 kb apart on the GAS chromosome; as with *emm* and *emm*-like genes, transcription of *sof* is positively regulated by the Mga regulon (78, 79). *sof* types have been defined based on ~600 to 750 bp positioned near the 5′ end of the *sof* gene (80). However, *sof* typing is far less developed than *emm* typing, and attempts to generate a well-supported phylogeny for the *sof* type-specific determinants have largely failed due to extensive intragenic recombination (81).

The relationships between the presence or absence of SOF activity, or the *sof* gene, and the *emm* pattern and *emm* clade as inferred from the *emm* type, are summarized in Table 4. Based on the strong associations between the MAP II material, class II M proteins, and *emm* pattern E (60), it is not surprising that there are strong correlations between *emm* pattern E and the presence of the *sof* gene and/or a positive serum opacification reaction (50, 80, 82, 83). Exceptions include several *sof*-positive strains within *emm* cluster E6 that are assigned to the *emm* pattern D grouping via PCR-based mapping (*emm59*, *emm81*, *emm85*); *emm12* isolates (pattern A-C) have an inactive *sof* gene. Despite the close genome map positions of *sof* and *emm*, a single *emm* type can be found in association with more than one *sof* type and vice versa (80), indicative of HGT of *emm* or *sof* to new genetic backgrounds and intraspecies recombination within the greater *emm* region.

T-Antigens

T-typing is based on agglutination of trypsin-treated GAS cells with hyperimmune rabbit serum (84, 85). The T-typing antisera is produced by immunization with

TABLE 4 Relationships between SOF/*sof* and *emm* patterns and clades[a]

emm type-based grouping	No. of *emm* types	% SOF activity positive	% *sof* gene positive	SOF/*sof* positive *emm* types	SOF/*sof* negative *emm* types
Pattern A–C	22	0	5.6	12[b]	All others
Pattern D	42	7.3	7.1	59, 81, 85	All others
Pattern E	48	97.9	97.9	All others	15
Clade X	57	87.7	87.7	All others	15, 34, 42, 51, 65, 67, 99
Clade Y	56	0	1.8	12	All others

[a]Based on analysis of 113 *emm* types, as reported in reference 83.
[b]The SOF product of *sof12* is typically inactive.

trypsin-treated GAS cells, followed by adsorption of antisera to undigested GAS cells of different T types; ~22 T types have been serologically defined. The T6 antigen gene (*tee6*) was the first one cloned (86) and was subsequently mapped to a region of the GAS genome that had been previously characterized as having genes encoding microbial surface components recognizing adhesive matrix molecules that bind fibronectin and collagen (87–89). The acronym FCT (fibronectin-binding, collagen-binding, T-antigen) is widely used to describe this genome region (87). SDSE isolates also express similar T-antigens (90), as do group B streptococci (91, 92).

T-antigens are equivalent to pili. Within the FCT region lie several genes that contribute to the production of pili whose assembly is mediated by sortases (48) (reviewed in references 93–95). Seven FCT region forms are depicted in Fig. 3 (82, 92); the FCT-3 and FCT-4 regions are associated with the vast majority of *emm* types (82). For the most prevalent FCT region forms (i.e., FCT-1, 3, 4, and 5) found among GAS strains of different *emm* types, the pilin backbone (BP) proteins and many of the ancillary pilin (AP) proteins that act as adhesins display extensive sequence heterogeneity (87, 92, 96).

The relationships between T-typing sera and recombinant pilin proteins has been examined (48, 92, 97). In perhaps the most thorough study to date, using 15 T-typing sera, agglutination of trypsin-treated GAS cells corresponds most closely to reactivity with both the BP and AP1 (presumptive adhesin) subunits by immunoblot (92). For a T3/13/B3264 strain (Alab49; FCT-3), the recombinant AP1 subunit (i.e., Cpa) is strongly reactive by immunoblot with each of the anti-T3, anti-T13, and anti-B3264 sera, whereas the BP (i.e., FctA) is bound by anti-T3 sera only (97). However, there is very little amino acid sequence homology between Cpa and FctA, as is also the case for the other adhesin-backbone pilin subunit pairs (82, 98).

That some T-typing sera display immunoreactivity to multiple recombinant pilin proteins seems likely due to incomplete adsorption of the antisera and/or the presence of antigenic epitopes shared among pilin subunits belonging to the variety of GAS strains used for immunization. Possible unmasking of immunological determinants via the trypsin pretreatment used in the cell agglutination assay may further complicate efforts to draw direct connections between T-serotypes and specific pilin protein variant forms. Adding even further complexity to the T-typing scheme are the numerous GAS isolates that agglutinate with multiple T-typing sera, such as T3/13/B3264 or T8/25/Imp19; many other T-serotype combinations (>60 in total) are also observed (83). Overall, the T-typing sera displaying the least cross-reactivity with multiple pilin forms are T2, T4, and T6 (92); these appear to be best represented by the FCT-6, FCT-5, and FCT-1 regions, respectively (Fig. 3), although within these FCT regions lie at least some extensive pilin gene diversity (see, e.g., reference 99 for FCT-1).

T-serotype and M-serotype and/or *emm*-type data for GAS isolates ($n > 20{,}000$) from two large collections (the CDC and the University of Minnesota reference laboratory), recovered in the United States from about the 1980s to the mid-2000s (83), was collated and is summarized in Table 5. There are no strict sampling criteria for this assemblage; it seems reasonable to assume that it largely consists of a combination of isolates causing iGAS disease and strains with high prevalence. In this skewed, non-population-based sample, the dominant T types among *emm* pattern A-C strains are T1, T11 and/or T12, and T3, T13, and/or B3264, reflecting the highly prevalent M/*emm* types M1, M12, and M3, respectively.

T-non-typeable (TNT) GAS probably represent isolates that fail to produce pili; they account for ~9% of total isolates but >20% of *emm* pattern D isolates (Table 5). The relatively large numbers of TNT isolates among pattern D strains may reflect the disproportional presence of *nra* (82); Nra functions as a repressor or activator of pilus gene expression in a strain-dependent manner (88, 100). Nearly 60% of *emm* pattern D isolates are T3, T13, and/or B3264 (Table 5), which

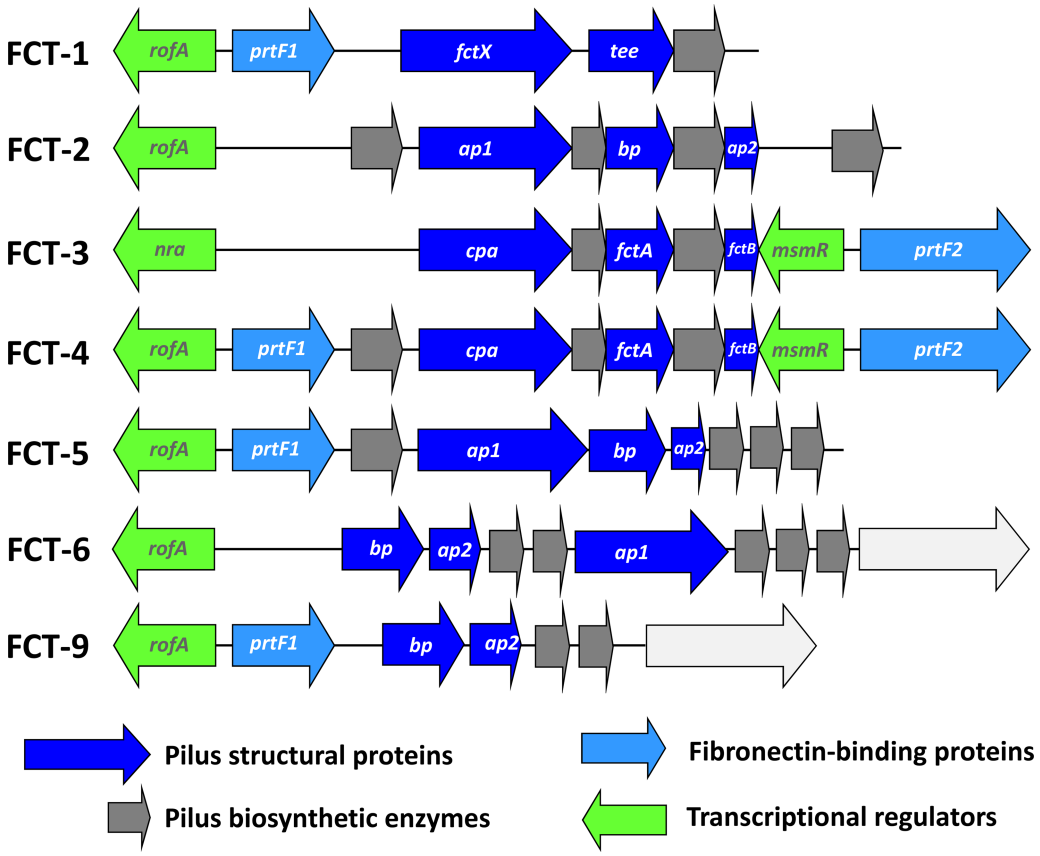


FIGURE 3 Structure of the pilus-encoding FCT region. Seven forms of the FCT region are shown. For pilus structural protein genes: *bp*, *tee*, and *fctA*, backbone pilins; *ap1*, *ap2*, *fctX*, *cpa*, and *fctB*, ancillary pilins. In FCT-2, *ap1*, *bp*, and *ap2* have sequence homology with *cpa*, *fctA*, and *fctB*, respectively, although divergence is high (82). Pilin adhesins include FctX from FCT-1, AP1 from FCT-2, and Cpa from FCT-3 and FCT-4. Pilin subunits that anchor the pilus shaft to the cell wall include AP2 and FctB (249). Fibronectin-binding protein genes include *prtF1* (sometimes designated *sfbI*), *prtF2* (sometimes designated *pfbp*), and possibly others that remain to be classified.

appears to be largely represented by the FCT-3 region form (82, 87, 92, 97). The most frequent M/*emm* types among pattern E isolates of this collection are M4 and M28 (83), which largely reflects the dominance of the T4 and/or T28 serotype grouping (Table 5). Several correlations between individual loci within the FCT region and *emm* pattern and/or *emm* type have been noted (50, 82, 92, 101).

Within an *emm* cluster group, there can be a very wide array of T types (based on Table 5; detailed data not shown) (83). For example, the *emm* cluster E3 isolates (13 *emm* types) are found in association with T5, T27, and/or T44 (28%); TNT (16%); T8, T25, and/or Imp19 (11%); T14 (11%); T9 (9%); T4 and/or T28 (9%); T3, T13, and/or B3264 (6%); and T11 and/or T12 (4%), among others. In contrast, cluster D4 isolates (21 *emm* types, all encoding plasminogen-binding M proteins) are far more homogeneous, with 72% associated with T3, T13, and/or B3264 (another 16% are TNT). Although *emm* genes within a cluster are more closely related to each another than to *emm* of other clusters, the extent to which surface pili contribute to the overall antigenic cell surface repertoire appears to vary widely by cluster.

For GAS sharing the same *emm* type, individual isolates often differ in their T-serotype (80, 83). For example, while a substantial majority of M/*emm* type 4 isolates from the Table 5 study are T4, there are other M/*emm4* isolates that have T types T4/28, T8/25/Imp19, T3/13, and TNT (83) (detailed data not shown). M/*emm4* isolates typically have *sof4*, but there are exceptions here as well (80). The unique combinations of M/*emm*, SOF/*sof*, and T types is indicative of HGT of the epidemiological marker genes among GAS.

TABLE 5 Relationships between *emm* pattern and T types for >20,000 GAS isolates[a]

T type	% Total	% Pattern A–C	% Pattern D	% Pattern E
T1	18.73	34.64	0.93	0.14
T2	3.30	0.00	0.00	7.85
T6	6.19	11.09	1.97	0.34
T9	0.82	0.00	1.39	1.81
T14	0.88	0.11	4.29	1.55
T15	0.00	0.00	0.00	0.00
T18	0.59	1.10	0.00	0.00
T22	0.02	0.00	0.00	0.04
T23	0.07	0.12	0.00	0.00
T3, 13, and/or B3264	16.10	17.14	58.63	10.75
T4 and/or 28	15.88	0.15	3.59	37.24
T5, 27, and/or 44	4.36	4.90	0.70	4.03
T8, 25, and/or Imp19	6.60	1.13	2.90	13.82
T11 and/or 12	18.17	20.80	4.52	16.17
Nontypeable	8.29	8.83	21.09	6.25

[a]Based on analysis of T-serotyping data (83) according to *emm* pattern groups or clades inferred from M or *emm* type (65, 126, 127), for 21,231 isolates, whereby 271 isolates had >1 grouping and were counted more than once. M or *emm* types ranged from types 1 through 124.

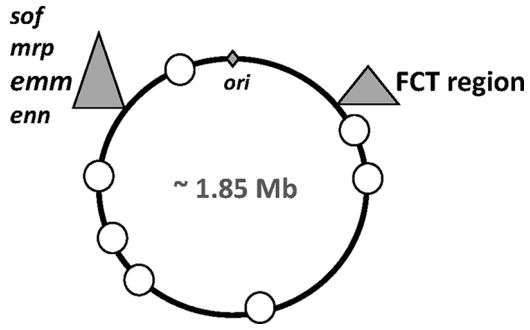
Unlike M type-specific antibodies, there is no known human host protective effect mediated via T type-specific serum; this may be a consequence of immunoreactivity directed toward trypsin-treated target antigens that are not encountered during natural infection. However, antiserum raised to recombinant pilin subunits or intact pili can be protective in experimental assays (48, 102). Shortcomings of the T-typing cell agglutination assay include a general lack of good resolution, the time-consuming process, and the requirement for standardized rabbit serum. Because T-typing sera display high levels of immunological cross-reactivity to different recombinant pilin polypeptides, and many GAS isolates agglutinate with multiple T-typing sera, it can be difficult to predict the FCT region genotype form based solely on the T-agglutination pattern.

At present, there is no single widely recognized genotyping scheme for the FCT region form and pilin genes (35, 82, 92, 96), but the need for one seems quite compelling given the importance of the FCT region in strain diversity as well as virulence.

Multilocus Sequence Typing and Core Housekeeping Genes

The first study to examine the population structure of GAS utilized multilocus enzyme electrophoresis (103). Isolates recovered from cases of severe iGAS disease correspond to several distinct lineages as defined by electrophoretic type. The multilocus enzyme electrophoresis approach was replaced by the more standardized multilocus sequence typing (MLST). For GAS, MLST is based on the nucleotide sequences of seven (partial) core housekeeping genes (104). MLST data are maintained at a user-interactive website (www.pubmlst.org/spyogenes) and currently lists 966 distinct sequence types (STs) (105); numerous investigators from throughout the world have generously contributed to this rich data set. Using goeBURST (106), 133 clonal complexes (CC) as defined by single locus variants (SLVs) have been detected. Figure 4 provides a chromosomal map of the relative positions of the core housekeeping genes and the *emm* and FCT regions.

FIGURE 4 Chromosomal map of the epidemiological marker genes of GAS. Open circles represent the seven core housekeeping genes used in MLST (104). The *sof* locus, when present, lies ~15 kb upstream of *emm*. The FCT region, which encodes pilus structural proteins and pilus biosynthetic enzymes plus other adhesins and transcriptional regulators, ranges in size from ~11 to 16 kb. The FCT region lies ~250 to 300 kb from the *emm* region, on the opposite side of the origin of replication (*ori*). GAS genomes range in size from ~1.8 to 1.9 Mb.



A given pattern A-C *emm* type has a high tendency to be associated with a single ST or CC (104, 107, 108). However, the relationships between *emm* type and ST are not strict, and *emm* types of pattern groups D and E are often found in association with distant STs (see "Lateral Movement of *emm*" below). These data parallel findings on the complex relationships between M/*emm* types, *sof* types, and T types (80, 83) and, taken together, provide supporting evidence for the HGT of *emm* from one genetic background to another.

Whole-Genome Sequences

The first completely assembled GAS genome was that of an *emm1* isolate (109). In the years that followed, numerous complete GAS genomes were reported for many of the most prevalent *emm* types found in resource-rich nations (updated listings are presented in recent reviews [50, 110]). Via next-generation sequencing technologies, many thousands of genomes of GAS isolates have undergone nucleotide sequence determination (e.g., 24, 30, 111–115). In nearly all studies, core genome sequences are compared for isolates sharing the same *emm* type, and numerous key insights on genetic changes that underlie enhanced transmission and/or virulence have been uncovered. The core genome data can be used to build evolutionary models of outbreaks and epidemics (see "Epidemiology: Building Evolutionary Models" and "Evolution: Short-Term" below).

A core genome MLST has not yet been developed for GAS. Although it is a reasonable concept for some bacterial species, its utility for GAS would probably be rather limited due to the ultrafine resolution this approach would provide for GAS, coupled with the extensive genetic recombination that characterizes this species. Core genome-based phylogenies of GAS of different *emm* types and/or CCs tend to yield lineages with deep branches, indicative of large genetic distances (50, 110, 111, 116).

The use of genome sequence data for surveillance of GAS has begun (35, 117–120). The CDC recently implemented a whole-genome sequence-based typing scheme for population-based iGAS strain surveillance conducted through the Active Bacterial Core program (35). For large-scale strain surveillance, the ability to obtain detailed information on typing and resistance features within a single output, from a single raw data input source, has largely precluded the need for multiple typing protocols and antibiotic susceptibility testing. The GAS typing pipeline automatically restricts an *emm* type to the sequence linked to *emm* consensus primer 1 (121) to improve accurate identification of *emm* and assigns an *emm* subtype (i.e., allele) as well. Partial sequences of BP genes and *sof* are also captured. In addition to *emm* subtype, the *emm*-like genes and MLST alleles and STs are identified, as is *gacI*, which is specific to CHO-A biosynthesis. Among numerous other genes, antimicrobial resistance determinants are also assessed.

A key feature of genome sequence-based active iGAS surveillance is the capacity to detect temporally and geographically related GAS clusters. For GAS isolates that are nearly identical at the single nucleotide polymorphism (SNP) level, the circumstances of where and when the isolates were recovered can be examined to help establish whether there exist disease clusters that may be well suited for an intervention (e.g., health care facilities, vulnerable patient populations such as the homeless, and IVDUs). Thus, next-generation sequencing coupled with an appropriate bioinformatics pipeline can be used as an autodetection tool to help guide public health interventions.

EPIDEMIOLOGY: GLOBAL TRENDS

Infection at the Epithelium of the Throat and Skin

Past studies show that GAS strains bearing certain M types have a strong tendency to cause pharyngitis but not impetigo. Similarly, there are M types often recovered from impetigo skin lesions but rarely from cases of tonsillopharyngitis (8, 122, 123). This observation gave rise to the important concept of classical throat and skin strains that are largely distinct from one another in their epidemiological and disease associations.

There are strong correlations between *emm* pattern groups A-C and D and recovery from cases of pharyngitis versus impetigo, respectively (50, 124, 125). Since the *emm* type is highly predictive of the *emm* pattern group for nearly all *emm* types examined (126, 127), reasonable inferences can be made about *emm* pattern grouping based on knowledge of *emm* types. Such inferences were made on an expanded scale for >5,000 GAS isolates derived from cases of pharyngitis or impetigo in 29 population-based surveillance studies from throughout the world (Table 6). In 23 studies on pharyngitis, *emm* pattern A-C and D isolates constituted a mean average of 41% and 6%, respectively. In six studies of impetigo, the mean average for *emm* pattern A-C and D isolates was reversed, at 9% and 47%. For pharyngitis versus impetigo groups, the difference between the fraction of A-C and D isolates in each surveillance study is highly significant (Table 6). *emm* pattern E isolates account for almost equal fractions

TABLE 6 *emm* pattern assignments for GAS recovered in 29 population-based surveys for pharyngitis or impetigo[a]

Disease association	No. of studies	Mean average % pattern A–C	Mean average % pattern D	Mean average % pattern E	Paired t-test[b] (% pattern A–C versus D)	Paired t-test (% pattern A–C versus E)[c]	Paired t-test (% pattern D versus E)
Pharyngitis	23	40.5 ± 19.8	5.8 ± 12.0	53.7 ± 16.0	<0.001	N.S.	n/a
Impetigo	6	9.3 ± 6.7	47.4 ± 8.1	43.4 ± 7.0	<0.001	n/a	N.S.

[a]Based on references 50 and 131. The 29 studies meet the following criteria: (i) isolates are clearly defined for recovery from cases of pharyngitis (or tonsillitis) or impetigo, and (ii) at least 25 isolates can be assigned to an *emm* pattern group based on *emm* type.
[b]Paired t-tests are two-tailed.
[c]N.S., nonsignificant; n/a, not applicable.

of throat and skin infections; the mean average percentage of *emm* pattern E isolates shows no significant difference from pattern A-C isolates at the throat or pattern D isolates at the skin.

Based on the strong clinico-epidemiologic associations, the *emm* pattern A-C strains are designated "throat specialists," whereas the pattern D strains are "skin specialists." As a group, *emm* pattern E strains are considered "generalists." Importantly, the *emm* pattern serves as a useful genotypic marker for the preferred tissue site of infection by GAS (32, 50, 124, 125, 128).

The *emm* cluster typing scheme can also be applied to the 29 population-based surveillance studies on GAS pharyngitis and impetigo (Table 7). Both pharyngitis and impetigo isolates are almost evenly split between clades X and Y, reflected by fractional ratios of percent pharyngitis to percent impetigo approaching 1. However, within clade Y, the distribution of pharyngitis and impetigo isolates are highly skewed among the clusters, with clusters A-C1-5 disproportionately represented by pharyngitis isolates, and clusters D1-5 disproportionately represented by impetigo isolates.

Several predicted binding motifs within M protein strongly correlate with *emm* clusters (65). Based on the 29 population-based surveys, the fractional ratio of pharyngitis to impetigo isolates is highly skewed for the fibrinogen-binding clusters of clade Y (high ratio value, with pharyngitis dominant) and the plasminogen-binding cluster D4 (low ratio value, with impetigo dominant) (51). The differential binding properties for human host proteins, by M proteins in accordance with disease, are potential determinants of tissue tropism. How modulation of the coagulation-fibrinolytic pathways (129) may translate into GAS tissue tropisms is an active line of inquiry.

Invasive Disease

As detailed above ("Ecology: Host Tissues and Diseases"), iGAS disease is extremely rare relative to GAS infection at the two primary ecological niches. When iGAS disease does occur, it is largely a transmission dead end, with the possible exception of direct contact transmission via open cutaneous wounds among underserved adult populations. Thus, iGAS infections have minimal

TABLE 7 *emm* cluster and clade assignments for GAS recovered in 29 population-based surveys for pharyngitis or impetigo[a]

emm clade	*emm* cluster	No. of isolates	No. of pharyngitis isolates	No. of impetigo isolates	% of total pharyngitis	% of total impetigo	Fractional ratio of pharyngitis to impetigo[b]	Fractional ratio of impetigo to pharyngitis
ALL	ALL	5,439	4,674	765	100	100	n.a.	n.a.
X	ALL	2,804	2,433	371	52.1	48.5	1.07	0.93
Y	ALL	2,608	2,239	369	47.9	48.2	0.99	1.01
Outlier	ALL	27	2	25	0.04	3.3	n.d.	n.d.
X	E1-6	2,802	2,432	370	52.0	48.4	**1.08**	**0.93**
X	Singletons	2	1	1	0.02	0.1	n.d.	n.d.
Y	A-C1-5	1,849	1,817	32	38.9	4.2	**9.29**	**0.11**
Y	D1-5	344	49	295	1.1	38.6	**0.03**	**36.78**
Y	Singletons	415	373	42	8.0	5.5	1.45	0.69

[a]Based on references 50, 51, and 135 and Table 6.
[b]Calculated as the ratio of the percentage of total pharyngitis isolates to the percentage of the total impetigo isolates.
[c]n.a., not applicable; n.d., not determined.
Key ratios are highlighted (bold)

impact on shaping the long-term evolution of GAS. The occurrence of iGAS disease is strongly affected by host risk factors and to a lesser extent by the genotype of the infecting GAS strain.

In the United States and Europe, most iGAS disease depends on transmission of GAS via a respiratory route (Fig. 1). Whole-genome sequencing of hundreds of GAS isolates has provided important details on genetic changes that enhance transmission and/or invasiveness (see "Evolution: Short-Term" below). The invasive index value provides insight on whether a GAS strain has an inherent ability to better gain access to deep tissue relative to its overall prevalence in the host population (22, 130). The *emm* types associated with iGAS disease have been intensively studied and are well described (e.g., 131–133).

Nonbullous impetigo caused by GAS usually arises in young children in tropical or subtropical climates. GAS can also cause slightly deeper cutaneous infections that are considered "pyoderma," and while they are not considered part of the iGAS disease spectrum, these cutaneous infections are distinct from typical impetigo in terms of patient demographics (i.e., adult, temperate regions) and host risk factors (e.g., IVDU, homeless, military) (23, 24, 134–136). Of interest is the strong association between certain pattern D *emm* types (specifically, *emm59* and *emm81*) and recent infections and outbreaks of deeper cutaneous infections. The *emm59* and *emm81* types are notable for their grouping in cluster E6 and tend to harbor *sof* as well; conceivably, these are "super-specialist" strains containing a hybrid mix of virulence factors that shift the modes of transmission and GAS disease course. Most other GAS-causing skin infection outbreaks among IVDUs are *sof*-positive *emm* pattern E strains (137, 138).

Acute Rheumatic Fever

Acute rheumatic fever (ARF) follows an inadequately treated GAS infection of the URT in a genetically predisposed host by ~3 to 6 weeks and can lead to rheumatic heart disease (RHD) via autoimmune attack of the heart valves (139–143). With >500,000 annual deaths, half attributed to RHD, GAS ranks high among all infectious causes of morbidity and mortality. Recent estimates of the prevalence of RHD are ≥15 to 30 million cases worldwide (15, 144–146); RHD is the leading cause of mortality from acquired heart disease in individuals <50 years old (147) and of all cardiovascular disease in children in resource-poor nations (148). Today, ARF and RHD are largely diseases of poverty, with an occasional resurgence in disease elsewhere (149).

Early epidemiological studies of ARF outbreaks in institutional settings showed that only a subset of GAS strains trigger ARF (11, 55, 140, 150–156). Historically, "rheumatogenic" GAS strains were referred to by their M protein serotypes, with M3, M5, M6, and M18 being highly prevalent. The concept of "nonrheumatogenic" strains is exemplified by the RHD patients of Irvington House during the preantibiotic era, who experienced a GAS outbreak (M type 4) in the absence of recurrent ARF attacks (155). Of ten M types often associated with ARF (M types 1, 3, 5, 6, 14, 18, 19, 24, 27, and 29 [157]), nine are *emm* pattern A-C (one is pattern E); of five relatively common M types that are rarely associated with ARF (M types 2, 4, 12, 22, and 28 [157]), four are *emm* pattern E (one is *emm* pattern A-C) (126, 127).

ARF-associated isolates recovered in Hawaii differ from the classic rheumatogenic *emm* types (158). Atypical *emm* types are also found in tropical Australia, where the prevalence of RHD is among the highest in the world (159). Over time, it has become much less clear which *emm* types are best assigned the "rheumatogenic" label, even though it is widely accepted that not all GAS strains trigger ARF in a genetically susceptible host (150, 155).

The potential role of GAS skin infection in triggering ARF and RHD in certain settings (e.g., noninstitutional, tropical) has been an ongoing debate since the 1960s (160). For example, the indigenous population of Australia suffers from a high incidence of ARF and RHD, yet GAS is far more frequently recovered from impetigo lesions than from the throat. Moreover, the strains recovered from the URT in this host population often belong to the *emm* pattern D skin specialist group or pattern E generalists. The epidemiological co-occurrence of GAS skin infections and RHD supports the idea that impetigo may contribute to this chronic disease. More extensive studies are needed to better understand skin infections and their potential role in ARF and RHD. The possibility that the closely related species SDSE might also trigger episodes of ARF or worsen RHD deserves deeper consideration as well (161, 162).

EPIDEMIOLOGY: DIVERSITY WITHIN HOST COMMUNITIES

Within a human host community over a narrow time frame, numerous GAS strains are in circulation. For the 29 population-based collections of GAS recovered from cases of pharyngitis and impetigo, the number of *emm* types and the Simpsons diversity index was calculated

TABLE 8 *emm* type diversity within communities for the 29 population-based surveys for pharyngitis and impetigo[a]

Disease association	No. of studies	Mean average no. of *emm* types	Mean average *D* index[b]
Pharyngitis	23	17.4 ± 7.8	0.873 ± 0.065
Impetigo	6	33.8 ± 8.1	0.959 ± 0.012
	Unpaired t-test, 2-tailed	$t = 0.0049$	$t < 0.001$

[a]Based on references50, 51, and 135 and Tables 6 and 7.
[b]Simpson's diversity index (*D*) (248) is based on *emm* types.

(Table 8). The number of *emm* types for the impetigo collections averaged nearly twice that of the pharyngitis collections ($t < 0.01$, unpaired t-test, 2-tailed). The Simpson's diversity index (*D*), which incorporates the number of recovered organisms into the calculation, also shows highly significant differences between the impetigo and pharyngitis collections, wherein the impetigo group has a higher mean average *D* value (a *D* value of 1 reflects maximal diversity). An analysis that includes all GAS, irrespective of clinical association, also shows fewer numbers of *emm* types accounting for a large proportion of isolates in resource-rich countries, where URT infections tend to predominate (131). The large numbers of circulating *emm* types is likely the consequence of expansive GAS strain migration, even into geographically remote communities (108). The biological basis for the difference in diversity among GAS in host populations experiencing pharyngitis versus impetigo may be tied to the mechanisms for transmission: respiratory droplets versus direct contact. It may be that the respiratory route is more highly efficient, such that a single infected individual can infect many more new hosts (i.e., higher basic reproductive rate), and this in turn leads to dominance by fewer strains, resulting in a lower *D* value.

EPIDEMIOLOGY: EVOLUTION OF MARKERS (*emm*)

Although *emm* type diversity within a host community is relatively low for the *emm* pattern A-C throat specialists compared to the pattern D skin specialists (Table 8), there is relatively high diversity among *emm* pattern A-C organisms in terms of the number of *emm* subtypes (i.e., alleles based on the 5′ end of the *emm* gene) that evolved worldwide (Table 9). The average mean number of *emm* subtypes identified for pattern A-C strains is 22.3 subtypes per *emm* type, compared to only 4.0 and 6.9 for pattern D and E strains, respectively, based on current data available in the CDC online database (58), which receives numerous deposits from investigators throughout the world.

The >5-fold difference in the number of *emm* subtypes for throat specialists versus skin specialists may be partly due to sampling bias that skews in favor of GAS invasive and pharyngitis isolates in resource-rich nations, such as the highly prevalent *emm1* type (pattern A-C). In addition, the tandem sequence repeats that overlap the *emm* subtype-determining region of *emm5* and *emm6* genes (pattern A-C) are a mutational hotspot due to DNA slipped-strand mispairing (163). However, biological factors such as host immune selection promoting immune escape (164) may also play a role.

Nucleotide sequence alignment of *emm* subtype alleles for each *emm* type-specific region, derived from >500 global GAS isolates ($n = 105$ *emm* types), reveals average ratios of nonsynonymous substitutions per nonsynonymous site (Ka) and synonymous substitutions per synonymous site (Ks) of 4.9, 1.5, and 1.3 for *emm* types of the *emm* pattern A-C, D, and E groups, respectively (107). Thus, diversifying (positive) selection acting on the *emm* type-specific sequence appears to be strongest for *emm* pattern A-C strains.

A likely source of diversifying selection acting on the *emm* type region is the protective host immune response, which can drive immune escape. Conceivably, the immunological pressures at the oropharynx (pattern A-C throat specialists) are more intense than those at the skin (pattern D skin specialists), although an underlying molecular mechanism is not presently obvious. Perhaps herd immunity builds up more quickly with a respiratory mode of transmission compared to direct skin contact; as herd immunity builds, the number of susceptible hosts dwindles, leading to more frequent occurrences of immune escape. In further support of a role for herd immunity in shaping the *emm* type content of GAS populations, newly acquired URT infections in older children are skewed toward less common *emm* types, consistent with exposure and protective immunity to the more prevalent *emm* types in early childhood (165).

The collection of M protein types assigned to a single *emm* cluster have high sequence similarity relative to M proteins of other clusters (65). Indeed, an M type-specific region sharing the same cluster as other M types

TABLE 9 Relationships between *emm* pattern or cluster and *emm* type and subtype

emm grouping category	Name of group	No. of *emm* types represented	No. of *emm* subtypes represented[a]	Average no. of *emm* subtypes per *emm* type
Pattern	A-C	35	781	22.31
Pattern	D	64	253	3.95
Pattern	E	67	460	6.85
Clade	X	85	518	6.09
Clade	Y	89	1,014	11.39
Cluster	E1	5	40	8.0
Cluster	E2	15	84	5.6
Cluster	E3	19	119	6.3
cluster	E4	17	150	8.3
Cluster	E5	7	15	2.1
Cluster	E6	18	98	5.4
Cluster	Clade X singletons	4	12	3.0
Cluster	A-C1	2	7	3.5
Cluster	A-C2	2	26	13.0
Cluster	A-C3	5	118	23.6
Cluster	A-C4	5	111	22.2
Cluster	A-C5	3	140	46.7
Cluster	D1	3	13	4.3
Cluster	D2	5	22	4.4
Cluster	D3	2	4	2.0
Cluster	D4	31	144	4.6
Cluster	D5	3	6	2.0
Cluster	Clade Y singletons	28	423	15.1
Clade Y singletons	Pattern A-C	16	385	24.1
Clade Y singletons	Pattern D	8	16	2.0
Clade X singletons	Pattern E	2	6	3.0

[a]Based on the CDC online database (58).

often elicits cross-protective antibodies (see "Epidemiology: Strategies for M Protein-Based Vaccines" below). Ultimately, this cross-protective effect may aid in the design of highly efficacious vaccines that have broad coverage but contain only a limited selection of type-specific epitopes. Whereas 90% of pattern D and E *emm* types belong to 1 of the 16 *emm* clusters, nearly half of pattern A-C *emm* types are standalone and do not cluster with any other *emm* type, highlighting once again a distinct dynamic for the evolution of many pattern A-C *emm* types.

Another possible factor at play in shaping the evolution of *emm* type-specific sequences is functional constraints, such as those imposed by binding C4BP via the type-specific region of numerous M proteins (166–168). Whole GAS organisms that bind C4BP tend to harbor M types associated with *emm* pattern groups D and E (167). Thus, there may be limits to diversification within M proteins that bind C4BP for the GAS strains that utilize this virulence strategy to enhance survival. Also, potentially related to selection pressures acting on GAS is the finding that engineered mutants with *emm* gene deletions have a (near) complete loss in the antiphagocytic effect for pattern A-C strains, whereas the loss is only partial for pattern D and E strains (169).

Overall, the throat specialists are characterized by fewer distinct *emm* types within a human host community but high diversification worldwide within the type-specific region (i.e., more *emm* subtypes). The skin specialists have about twice the number of *emm* types within a community, but diversification within the type-specific region is ~5-fold lower based on the number of *emm* subtypes or Ka/Ks ratios. The *emm* pattern E generalists are most similar to the skin specialists in these values.

The "*emm* type" and "*emm* cluster" definitions are categorical, and the noted differences between throat and skin specialists may partly reflect the placement of the categorical "cutoff" value along a continuum of sequence divergence. Factors that shape the GAS popula-

tion structure, as defined by *emm* type, likely include the route of transmission, environmental risk factors, the host immune response, tissue-specific host factors, binding functions of M protein, and intrinsic GAS factors that influence mechanisms for genetic change.

EPIDEMIOLOGY: STRATEGIES FOR M PROTEIN-BASED VACCINES

Numerous vaccine strategies for protection against GAS infections were recently reviewed (170–172). Briefly, vaccine strategies can be divided into M protein and non-M protein target antigens, with the latter including streptolysin O, C5a peptidase, CHO-A, pili, and the enzyme arginine deiminase (102, 173–178). However, only the vaccine candidates that incorporate M protein as the target antigen have attained human clinical trial status (179).

M protein vaccines based on the conserved antigens of the CRR (Fig. 2A) have undergone extensive investigation (49, 180–185). Since all M proteins have a CRR, a CRR-based vaccine has the theoretical advantage of providing broad coverage against all GAS strains; several formulations are currently, or will soon be, in clinical trials. Potential drawbacks of a CRR-based vaccine include immune escape. The CRR region of M protein, which includes MAP I/II and the class I/II antigenic epitopes, is not 100% conserved. For example, among 175 *emm* types from a worldwide strain collection, there are 22 alleles of the CRR-based J8 vaccine candidate (65). Importantly, some of the amino acid residues within the CRR of the M protein display signatures of diversifying selection. Immune escape is a possibility for any vaccine target, and current knowledge indicates that it deserves careful consideration for a CRR-based vaccine.

The second category of M protein-based vaccine candidates targets the NH_2-terminal type-specific determinants (179, 186). Pioneering work in the 1940s and 1950s showed that type-specific serum is responsible for protective immunity against the homologous *emm* type (187–190); in contrast, minimal to no protective effects were observed for infection by heterologous GAS types. A 26-valent M type-specific recombinant polypeptide vaccine is the only GAS vaccine to have reached phase II clinical trials, whereupon it was found to be safe and highly immunogenic (179). The "type-specific protection" paradigm of GAS immunity has led to concerns about the potential for low coverage by the multivalent vaccine, especially in tropical settings, where the number of GAS *emm* types can be very high (131, 191–195).

Recent data indicate that cross-protective immunity between closely related *emm* types can induce broader protection via multivalent vaccines than type-specific immunity might otherwise predict. An initial hypothesis on cross-protective immunity was raised following genetic analysis of GAS strains collected in Brazil (196, 197). Cross-protective opsonization by hyperimmune rabbit antiserum generated via immunization with M type-based vaccines was subsequently demonstrated using a 30-valent vaccine (186) and an experimental 4-valent vaccine designed to protect against the 17 *emm* types belonging to *emm* cluster E4 (198).

Superficial skin infection can stimulate a robust antibody response. A recent study of humoral immunity acquired after impetigo infection in Fijian schoolchildren demonstrates that 38% of subjects had a ≥4-fold increase in anti-M-type-specific IgG titers that match the M type of the infecting strain (199). Importantly, cross-protective *emm* cluster-specific immunity was evident for clusters E4, E6, and D4, although the levels of cross-protection were variable depending on both the patient and *emm* cluster. That the cross-reactive immune responses frequently align with *emm* cluster groups holds promise for the design of M type-based vaccines having broad coverage. More data are needed to further solidify the clinical and experimental findings obtained to date; identifying immunoassays to help define correlates of protection (CoPs) is a high-priority goal (200). Serotype replacement may potentially be an issue for GAS following widespread immunization with M type-based vaccines, as is the case for *Streptococcus pneumoniae* and serotype-based capsule vaccines (201, 202).

EPIDEMIOLOGY: BUILDING EVOLUTIONARY MODELS

MLST data can be used to build models of evolutionary descent via eBURST or goeBURST, which are clustering algorithms based on SLVs (and/or double locus variants) among the seven core housekeeping genes (106, 203). The single mismatched allele that defines an SLV pair can arise by either point mutation or homologous recombination, whereby the latter may introduce >1 SNP.

For GAS, the combined total length of the seven housekeeping (partial) alleles is 3,134 bp. However, 3,134 bp represents only a fraction of the GAS core genome; for example, *emm3* type iGAS have a core genome of ~1,645 kb (115), which is ~524-fold longer in sequence than the MLST alleles combined. Based on these values, it can be estimated that an MLST-defined SLV pair of GAS would differ by an average minimum

of >500 SNPs within its core genome if genetic change were strictly due to point mutation or intraspecies recombination between very closely related organisms. However, if there was extensive HGT and recombination between more distally related GAS, the number of SNPs that differentiate the SLV pair is expected to be even higher. Given that the core genome of *emm3* type iGAS strains has an estimated 29 to 133 SNPs per isolate (i.e., short-term evolution) (115), the models for evolutionary descent based on the SLVs of MLST using only seven core housekeeping genes may best capture intermediate stages of evolution, which can be quite informative.

MLST data in combination with epidemiological marker genotypes can be compared to whole-genome sequence data. A good example for comparison is the *emm89* strains causing iGAS disease in Europe and the United States, which underwent a marked increase in prevalence during the 2010s (30, 31, 111, 204). Using the ST and GAS isolate data deposited at www.pubmlst.org/spyogenes (105), a goeBURST plot that includes the recent epidemic *emm89* strains is shown in Fig. 5A. The newly emerged epidemic clade of *emm89* corresponds to ST101 (and its descendants). Based on demographic information (i.e., dates of recovery), it seems reasonable to assume that ST407 is ancestral to ST101. Notably, ST407 is recovered in association with two distinct *emm* types: at least one other ST407 GAS isolate in the PubMLST database has *emm78* in place of *emm89*, as do several other SLVs and double locus variants of ST407. Alignment of the 180-bp *emm* subtype regions of *emm78.0* (ST253) and *emm89.0* (ST407, ST101) reveals <60% identity for both nucleotide and amino acid sequences, indicative of substantial divergence (ClustalW; data not shown) and supporting a model wherein *emm78* and/or *emm89* underwent HGT and recombination onto an ST407-related genetic background. Both *emm78* and *emm89* are *emm* pattern E but belong to different *emm* clusters (E1 and E4, respectively).

Examination of the FCT regions of ST407 and ST101 *emm89* type GAS isolates reveals highly divergent Cpa and FctA proteins, corresponding to the AP1 and BP subunits of pili, respectively (Fig. 5B). The distinct Cpa and FctA forms appear to coincide with T serotypes T11 (ST407) and T13 (ST101) (35). The housekeeping gene locus that distinguishes ST407 and ST101 is *yqiL*, which lies adjacent to the FCT region (Fig. 4); *yqiL01* (ST401) differs from *yqiL03* (ST101) by three nucleotide substitutions, highly suggestive of diversification via recombination. In contrast to the eBURST analysis of MLST data, whole-genome sequencing and analysis of the core genomes of *emm89* isolates provides a much higher level of resolution (30, 35, 111, 112). Importantly, several SNPs and indels that distinguish the *emm89* clade 2 → clade 3 transition lead to alteration of virulence factor expression: loss of the hyaluronic acid capsule (Δ*hasA*) and transcriptional upregulation of toxin genes encoding Nga and streptolysin O.

The genetic relationships among GAS strains harboring *emm89* and ~20 additional *emm* types were inferred based on ~75,000 SNPs of the core genome (111). The phylogeny shows that the *emm89* isolates of the three clades detailed above, and corresponding to CC407 (ST101, ST407), cluster within the same major lineage; ST142 isolates (double locus variants of ST407; Fig. 5A) also cluster on that same deep branch. However, GAS of other *emm* types, plus *emm89*-ST380 isolates belonging to another CC (204), each cluster along distinct lineages with deep branches (111). ST380 differs from both ST101 and ST407 at all seven housekeeping alleles, indicative of HGT of *emm89* to a distant genetic background.

The MLST data for *emm89* strains captures two different modalities of HGT involving *emm*: replacement of a nascent *emm* by a new *emm* type on the original genetic background (i.e., recipient-new clone pairs, with two *emm* types and one ST) and lateral movement of an *emm* type to a new genetic background (i.e., donor-new clone pairs, with one *emm* type and two STs) (107). Within CC407 (recipient) there is evidence for the former process, involving replacement of *emm78* with *emm89* or vice versa (Fig. 5A). A possible donor of *emm89* to CC407 is *emm89*-ST380 (latter scenario); alternatively, ST380 may have been the recipient of *emm89* from a CC407 donor. Comparisons of completely assembled genomes of possible donor-recipient-new clone pairings may aid in establishing the directionality of horizontal *emm* gene movements.

Based on the connections within CC407, the seven housekeeping genes of MLST appear to recapitulate the more expansive core genome sequence data reasonably well. However, MLST fails to capture SNPs arising in core virulence-related genes that may alter the trajectory of GAS disease in a host population. What MLST may be best tuned to, when combined with typing data on *emm*, pilin genes, and/or *sof*, are HGT events wherein a typing gene is present on distant STs/CCs or multiple typing genes are recovered in association with the same ST/CC (such as divergent Cpa-FctA in *emm89*-CC407; Fig. 5B). For GAS, MLST relationships can provide an intermediary link between short- and long-term evolution and may help delineate the genetic roots of the emer-

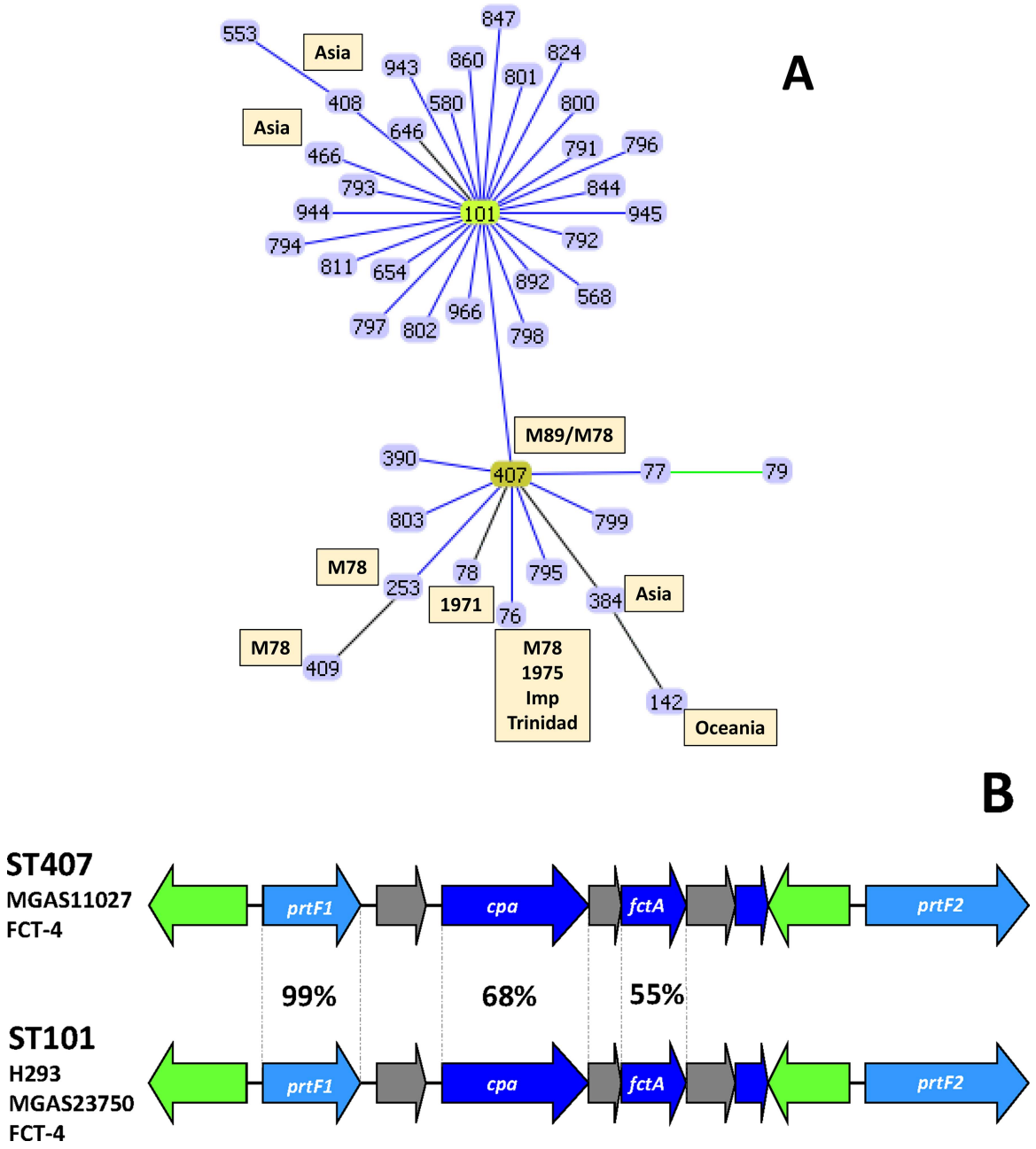


FIGURE 5 Model for evolutionary descent (CC407). **(A)** goeBURST diagram of CC407, which includes single representatives of 41 STs and is based on data from www.pubmlst.org/spyogenes (105). Unless noted otherwise (orange shaded boxes), all isolates (n = 114) are *emm89* and were recovered from (i) the 1990s through 2010s, (ii) Europe or North America, and (iii) URT infections or iGAS disease. For evolutionary analysis of *emm89* (formerly PT4245 [250]) isolates based on the core genome sequence (31, 111), clade 1 strain MGAS11027 is ST407, whereas clade 2 and epidemic clade 3 strains (MGAS23530 and MGAS27061, respectively) are ST101; the epidemic *emm89* strain H293 from the United Kingdom is also ST101 (30). Note that ST407 corresponds to GAS with either *emm89* or *emm78* (105). **(B)** Amino acid sequence identities among predicted FCT region gene products of ST407 (MGAS11027) and ST101 (H293, MGAS23530) strains (all *emm89*) via ClustalW alignments.

gence of new strains that have unique repertoires of cell surface antigens and their linked virulence functions.

EVOLUTION: MOLECULAR MECHANISMS OF GENETIC CHANGE

Core genes of GAS contain a historical record of numerous homologous recombination events (111, 205–208). Intraspecific gene exchange between GAS strains most likely occurs during cocolonization of the URT or coinfection of impetigo lesions with multiple strains (209). GAS can grow as biofilms or form microcolonies during infection within both tissue niches (210, 211), and the high bacterial density that is attained likely facilitates HGT. There is also ample evolutionary evidence for gene exchange between GAS and other species, such as SDSE (212–218).

A major unresolved puzzle of GAS biology is the molecular mechanisms that govern HGT, leading to the high rates of homologous recombination that is inferred from its evolutionary history. In principle, nonspecific packaging of chromosomal fragments into bacteriophage capsids could provide a vehicle for HGT of bacterial DNA (i.e., generalized transduction). In addition to bacteriophage, GAS are also rich in integrative and conjugative elements (91). The degree to which random DNA fragments may slip through during the conjugative process remains unclear. A conserved RNA uracil methyltransferase (*rum*) gene provides an integration hotspot for both bacteriophage and integrative and conjugative elements in GAS, as well as other related species (219). Natural transformation via high-frequency uptake of naked DNA has not been demonstrated for GAS, although competence gene homologues are present in the GAS genome and (very) low efficiency transformation is detectable in biofilms (220–222). However, GAS secrete numerous DNases that degrade extracellular DNA, which may render the transforming substance unavailable for uptake by a competent cell.

A mutator phenotype that results in an increased mutation rate has been characterized for GAS. This defect in DNA mismatch repair is controlled by excision and reintegration of a streptococcal phage-like chromosomal island (SpyCI) at a site between the *mutS* and *mutL* loci (223) and is present in the genomes of numerous GAS strains (50). SpyCI yields an ~100-fold increase in the rate of spontaneous mutation, to as high as $\sim 10^{-7}$ to 10^{-8} mutations per generation. A possible relationship between the presence of SpyCI and CCs with a relatively low ratio of recombinational to mutational events remains to be explored.

EVOLUTION: LONG TERM

Core Housekeeping Genes

As previously stated, core genome-based phylogenies of GAS of different *emm* types and/or CCs tend to yield lineages with deep branches, indicative of large genetic distances (50, 110, 111, 116). Whole-core genome phylogenies depicted as SplitsTree graphs (neighbor-net analysis) show recombination between GAS of different *emm* types located deep within the network. Furthermore, there is no obvious clustering of the major lineages in accordance with *emm* pattern grouping or by any other recognized parameter.

Numerous methods have been used to assess the relative amount of recombination versus mutation that contributes to genetic change in GAS, based on the seven core housekeeping genes used in MLST (205–208). Together, the findings support the notion that GAS exhibit relatively high levels of recombination which, at least by some measures, is roughly comparable to the levels observed for the pneumococcus.

Using MLST data, the relative amount of recombination versus mutation among the *emm* pattern-defined subgroups of GAS was examined (107, 224). Findings based on several methodologies all point to the same general trend, wherein the pattern A-C throat specialists have the highest tendency to diversify by mutation, and pattern D skin specialists and pattern E generalists have a greater tendency to diversify by recombination. The data on a larger role for mutation, relative to recombination, in genetic diversification among the classical throat strains is consistent with the molecular evolutionary analysis of *emm* subtypes (see "Epidemiology: Evolution of Markers (*emm*)" above; Table 9) (107). Differences among the *emm* pattern-defined groups of GAS in terms of their mechanisms for genetic change underscore different dynamics in shaping their long-term population structures.

GAS strains may encounter ecological separation due to their tissue site preferences for infection (throat versus skin), geographic partitioning (temperate versus tropical) and temporal distances (winter versus summer). Ecological barriers can give rise to allopatric speciation, and therefore, signatures of early stages of speciation within the GAS population may be evident. However, analyses of housekeeping alleles provide strong support for highly fluid HGT and homologous recombination of housekeeping genes among all three *emm* pattern groupings (101, 128, 224). These findings extend to isolates known to be recovered from the URT versus impetigo lesions, regardless of *emm* pattern group (224).

In summary, despite wide spatial-temporal distances between many GAS strains, there is no clear evidence for diminished opportunities for HGT of core housekeeping genes between organisms of the different *emm* pattern groups. Numerous statistical measures all point to a population genetic structure for the *S. pyogenes* species that is highly nonclonal (205–208). Against a background of highly random genetic change in core housekeeping genes, loci that exhibit strong linkage with *emm* pattern genotypes are reasonably good candidates for playing a direct role in the adaptations leading to throat versus skin infection (e.g., 50, 101).

Lateral Movement of *emm*

For the same global set of >500 GAS isolates used to calculate Ka/Ks ratios within the *emm* type-specific region (see "Epidemiology: Evolution of Markers (*emm*)" above), the number of HGT events as a function of *emm* type was measured (107). Movement of *emm* to a distal ST is quantified as recovery of an *emm* type in association with >1 ST differing by ≥5 of the 7 housekeeping alleles; data show an average of 0.17, 0.60, and 0.75 lateral transfer events per *emm* type for the pattern A-C, D, and E groups, respectively. This finding supports the notion that for the classical throat strains, but less so for skin strains or generalists, *emm* type is often a reasonably accurate marker for ST or CC (104, 107, 108). In another measure of HGT and recombination, the number of *emm* types associated with *emm* type-variable STs (i.e., STs recovered in association with >1 *emm* type) was calculated; the number of *emm* type replacements per ST was 0.12, 0.15, and 0.02 for pattern A-C, D, and E isolates, respectively (107). Overall, there were three times as many events captured for HGT of *emm* to a distant ST (n = 63) than for replacement of *emm* on the same ST (n = 21).

Taken together, these data shed light on GAS strain competition surrounding the narrow time-space wherein the donor, recipient, and newly emerged clone all coexist. Strain competition can be mediated through host/herd immunity (225) and perhaps via other host-pathogen interactions as well. For the throat specialist strains, competition appears to be most intense between strains sharing the same *emm* type, because organisms of the original ST that served as the (presumptive) *emm* donor are rarely recovered. In contrast, for generalists, competition at the level of *emm* type appears to be considerably weaker, as evidenced by the frequent recovery of the same *emm* type in association with multiple distant ST backgrounds. How other cell surface antigens, such as SOF and/or pili, might contribute to the competitive sphere will require study of lateral movements of their genes and alleles within the GAS population.

Full Complement of Epidemiological Markers Considered

How do the host's specific immune defenses perceive the incoming antigenic repertoire present on the GAS cell surface? Except during carriage, the Mga regulon that upregulates the expression of *emm* and *emm*-like genes and *sof* is typically active. The pilus T-antigens appear to be expressed by >90% of the GAS population. The expression/nonexpression of the hyaluronic acid capsule can potentially mask/unmask some or all the GAS cell surface proteins. Setting aside the variance in the transcriptional program, which can have a profound effect on transmission and virulence (see "Evolution: Short Term" below), the collection of epidemiological markers provides a useful construct for defining GAS "strains."

Figure 6 summarizes the relative diversity or presence/absence of each major epidemiological marker in accordance with the *emm* pattern-defined groupings of throat specialists, skin specialists, and generalists. Importantly, *emm* type is rarely recovered in association with multiple *emm* pattern groupings (Table 2), thereby allowing reasonably valid inferences to be drawn. The pattern- and cluster-defined *emm* subtype alleles reveal stark differences between the throat specialist, skin specialist, and generalist groupings regarding diversification (Table 9). Major distinctions also lie in the relative distributions of the *sof* locus (Table 4), *emm*-like genes (see Fig. 2B), and FCT region forms (Fig. 3) (82).

Do highly diverse *emm* subtypes compensate for the dearth of *emm*-like genes and *sof* among the throat specialists—and vice versa for the generalists? The diversity among *mrp* and *enn* alleles is modest, but they too have undergone intraspecies HGT (226); selection may be tied to functional binding attributes of the Mrp or Enn proteins (e.g., IgA, IgG) (63, 227–230). The *sof* gene also undergoes rather extensive intraspecies HGT (80); however, *sof* gene movement is nonrandom in that it is largely restricted to the *emm* pattern E generalists (Table 4). Yet the relationships between individual *emm* types and *sof* types are highly varied, with many unique combinations arising via HGT to distant genetic backgrounds (35).

Among the throat specialist group is a wide variety of FCT region forms (Fig. 6) (82). The FCT-1 region is the most prevalent form among representative pattern A-C organisms of unique *emm* types; FCT-1 is characterized by an AP1 adhesin (FctX) and major BP (Tee) that are markedly divergent from the highly variable

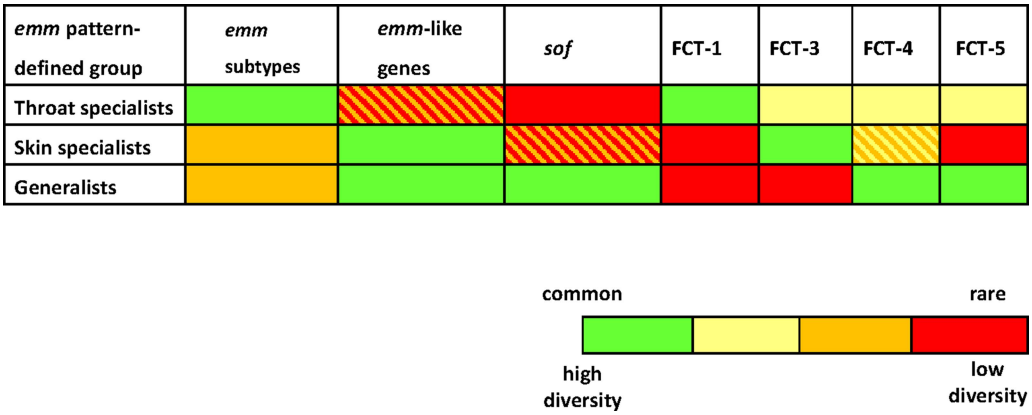


FIGURE 6 Trends in the distribution of epidemiological markers. An overview of the diversity and/or occurrence of epidemiological marker genes among three clinico-epidemiologic phenotype groupings of GAS. The relative distribution of FCT region forms is in accordance with one representative isolate for 94 unique *emm* types, as reported in reference 82; however, given the extensive HGT of *emm* and/or FCT genes to distant genetic backgrounds, the present analysis of the FCT region distribution may be far from complete.

Cpa and FctA pilins of FCT-3 and FCT-4 (82, 92). Not shown in Fig. 6 is FCT-2, which is (almost) exclusive to the highly prevalent *emm1* organisms; the AP1 and BP pilins of FCT-2 are akin to a distant lineage of Cpa and FctA, respectively (82). Thus, the unusually high success (i.e., fitness) of *emm1* strains may be due, in part, to its antigenically unique pilus structure. Overall, the absence of *sof* and *emm*-like genes among the pattern A-C strain grouping may be compensated for by the antigenic and structural diversity among its pili.

The skin specialist group appears to be the most homogeneous regarding epidemiological marker loci content, largely lacking *sof* but having both *mrp* and *enn*, and an FCT-3 region skewed heavily toward T3/13/B3264 (Table 5); this genotype is coupled to the lowest diversity among *emm* subtypes (Table 9). Also, the skin specialist group is dominated by *emm* cluster D4, encoding the plasminogen-binding M protein (51, 65). However, past lateral movement of *emm* type to distant genetic backgrounds is rather abundant among the skin specialists and is ~2-fold greater than that of *emm* type replacement on the same ST (107) (see "Lateral Movement of *emm*" above); there is also ample HGT of housekeeping alleles. Conceivably, the *emm* pattern D skin specialists represent a more recent subpopulation of GAS (i.e., younger) that has less diversity within any one gene, but the genes themselves are highly recombinatorial. Alternatively, the cell surface protein repertoire of skin specialists may be subject to less intense pressures for diversifying selection, perhaps due to its microenvironment and/or the host immune response at the skin.

The relative lack of pattern D and E *emm* types recovered in association with FCT-1 or FCT-2 regions may be indicative of functional constraints that would otherwise lower fitness, such as inadvertent adherence to a suboptimal tissue via specific host receptors. As a group, the *emm* pattern E generalists display all the *emm* region epidemiological markers and tend to have either the large FCT-4 region or FCT-5 (Fig. 6), both of which harbor the *prtF1* gene that encodes a fibronectin-binding protein (Fig. 3). Pattern E strains are also characterized by relatively high levels of recombination, involving both *emm* and the seven core housekeeping genes of MLST (50, 231, 232). Conceivably, within the generalist grouping are discrete throat or skin specialist strains whose phenotype is masked by the organizational framework (i.e., *emm* pattern group) that is imposed.

Historically GAS "strains" have been defined (primarily) by M serotype and secondarily by T and/or SOF serotype in combination with M type. Over the past 2 decades, *emm* typing has largely replaced GAS serotyping. The multitude of combinations of *emm* type and ST underscores the HGT of *emm* type, which is also consistent with the diverse array of M, T, and SOF serotype combinations observed. Even within a CC, STs are found in association with multiple *emm* types and/or divergent pilin genes (e.g., Fig. 5). Therein lies a conundrum for studying the population biology of GAS: How does one best define "strain?" This may or should be one of the next new challenges to be addressed by the GAS field.

EVOLUTION: SHORT-TERM

The monophyletic grouping of GAS strains in accordance with *emm* type and/or MLST-defined CC readily allows for inference of ancestral-descendent pairs within a branch, especially when coupled to epidemiological collections of strains with known dates of recovery. Whole-genome sequences of hundreds or thousands of GAS isolates sharing the same *emm* type and collected over a narrow spatial-temporal distance have captured epidemic waves of clones that are distinct from presumptive ancestral clones (24, 30, 111, 113–115, 233). An important finding is that within an *emm* type (specifically, *emm1* and *emm3*), iGAS isolates arise from all major lineages occupied by URT isolates and are genetically similar to the circulating pharyngitis strains from which they evolved (114, 234).

Comparative genomics of GAS recovered within a narrow spatial-temporal distance, and in association with a defined bacteria-host ecology (i.e., pharyngitis, asymptomatic carriage, and/or iGAS disease), have revealed high levels of diversifying selection at certain loci, particularly those involved in regulating the phenotypic expression of virulence factors (24, 30, 44, 111, 113–115, 233–240). Affected loci include *hasABC* (hyaluronic acid capsule biosynthesis), *covRS* (two-component system that acts as a global regulator), *rgg/ropB* (standalone response regulator), *mga* (response regulator and activator of *emm* transcription), *lia* (sensor kinase), and *nga-slo* (exotoxins). There are some recurrent themes, wherein downregulation of *emm* enhances URT carriage, downregulation of *speB* enhances iGAS disease, and upregulation of *nga-slo* enhances URT transmission. The contribution of capsule to GAS ecology seems more complex, and perhaps it is GAS strain dependent. In fact, in general terms, all mutations that alter fitness are context dependent and, accordingly, may be strain dependent. In addition to SNPs in critical virulence loci, GAS genomes are profoundly shaped by mobile genetic elements (which are often not included as part of the core genome); bacteriophage and integrative and conjugative elements can harbor critical virulence genes as well (e.g., 114).

EVOLUTION (OR NOT): RESURGENCE OF SCARLET FEVER

Scarlet fever (SF) was a dreaded GAS disease through the early 1900s; it manifests as a URT infection coupled to a pathognomonic scarlatina rash triggered by the action of erythrogenic toxin, which was subsequently characterized as a T-lymphocyte superantigen. The occurrence of SF in the United States and Europe declined mid-century and was largely sporadic in recent decades. However, in 2011 there was an alarming increase in the number of reported cases of SF in Hong Kong and in China soon thereafter (241–243). This upsurge was largely attributed to *emm12* GAS organisms that were resistant to macrolides and tetracycline. Comparative genome sequence analysis with other *emm12* strains revealed a novel prophage harboring the superantigen genes *speC* and *ssa* (113, 244). However, the Hong Kong SF outbreak was multiclonal in origin.

A resurgence of SF cases also occurred in the United Kingdom, beginning in ~2013 (245). Comparative genomics failed to reveal a single clone or a single gene (e.g., a superantigen gene on a mobile genetic element) associated with those disease cases. As in Asia, the increase in SF incidence was associated with multiple lineages of GAS (246, 247).

Taken together, the recent upsurge in SF appears to be multifactorial and likely includes host-related factors. Numerous GAS lineages are responsible for SF, although resistance to antibiotics may have fueled the spread of some clones. Possible shifts in the content of superantigen genes may collectively have had an effect. The two cohorts—children in Asia and Europe and largely nonrural—possibly had a history of past exposure to a similar set of prevalent *emm* types and strains (131), and their herd immunity may be similar despite the geographic distance. Conceivably, an unknown copathogen may affect host susceptibility to SF and development of the scarlatina rash. Heightened clinician awareness and better reporting and surveillance, and/or changes in clinical practice, can also contribute to a higher disease incidence. The mysterious underpinnings of the recent upsurge in SF cases in Asia and Europe underscore the contributory role of factors other than genetic changes in GAS in shaping the epidemiology of GAS disease.

EVOLUTION AND POPULATION DYNAMICS

A central problem concerning GAS is the difficulty in connecting the evolutionary histories of different *emm* type-defined strains. There are large genetic gaps between extant organisms—but also rampant gene flow. So, what became of the evolutionary intermediates? Was there an extended series of population bottlenecks whereby new clones outcompeted their progenitors? And perhaps each successful new clone bore a new *emm* subtype plus changes in other genes, until a very unique and highly fit organism with a far more divergent *emm*

type emerged and gained a strong foothold in the human host population. Perhaps the genetic intermediates went extinct due to low fitness in the context of shifting host-environmental conditions, or perhaps they remain extant, lurking in low numbers in a semiquiescent state and have yet to be sampled.

When did GAS emerge as a human pathogen? Was the GAS progenitor a human commensal? Or was the progenitor an animal pathogen that has yet to be recovered via sampling? Did the new GAS pathogen acquire key genes from a trading partner as a pivotal step in its emergence? What are the genotypes that confer adaptation of GAS to the primary ecological niches of the throat and skin? Are GAS commensals at heart, and should one view their ecology primarily through that lens?

What was the structure of the human population at the time GAS emerged as a human pathogen—highly migratory with intermingling or isolated tribes? What is the minimum host population size necessary to sustain GAS transmission? What becomes of a GAS strain if/when it runs out of susceptible hosts? Does the organism persist in an altered carrier state? Why do there exist ~243 *emm* types? And why are there even far larger numbers of unique combinations of M, SOF, and T antigens? Is the number of unique combinations of cell surface proteins related to characteristics of the human host population? If so, how?

What are the selective pressures shaping the GAS population of today? Is the ability to cause a clinically inapparent infection a more recent adaptation, driven by social behavior in the form of seeking antibiotics when ill, thereby giving the stealth invader a survival advantage? Why did SF recently re-emerge in Asia and the United Kingdom? Why did acute rheumatic fever undergo a renaissance in the United States during the 1980s? How might a future vaccine alter its evolutionary trajectory? A better understanding of the ecology and evolution of GAS, and its population genetic structure, may ultimately lead to interventions that limit its impact on human health.

ACKNOWLEDGMENTS

Funding support for this work is from the NIH (grant AI-117088 to D.E.B.) and FNRS research grants (PDR T.0255.16 and CDR J.0019.17 to P.R.S.).

The authors dedicate this work to Susan K. Hollingshead (1953-2017), a coauthor of earlier editions of this chapter and a pioneer in the field of group A streptococcal genetics and evolution.

REFERENCES

1. García A, Paul K, Beall B, McClure H. 2006. Toxic shock due to *Streptococcus pyogenes* in a rhesus monkey (*Macaca mulatta*). *J Am Assoc Lab Anim Sci* **45:**79–82.

2. Hook EW, Wagner RR, Lancefield RC. 1960. An epizootic in Swiss mice caused by a group A *Streptococcus*, newly designated type 50. *Am J Hyg* **72:**111–119.

3. Hysmith ND, Kaplan EL, Cleary PP, Johnson DR, Penfound TA, Dale JB. 2017. Prospective longitudinal analysis of immune responses in pediatric subjects after pharyngeal acquisition of group A streptococci. *J Pediatric Infect Dis Soc* **6:**187–196 http://dx.doi.org/10.1093/jpids/piw070.

4. Kaplan EL. 1980. The group A streptococcal upper respiratory tract carrier state: an enigma. *J Pediatr* **97:**337–345 http://dx.doi.org/10.1016/S0022-3476(80)80178-8.

5. Cleary PP, McLandsborough L, Ikeda L, Cue D, Krawczak J, Lam H. 1998. High-frequency intracellular infection and erythrogenic toxin A expression undergo phase variation in M1 group A streptococci. *Mol Microbiol* **28:**157–167 http://dx.doi.org/10.1046/j.1365-2958.1998.00786.x.

6. Fiedler T, Köller T, Kreikemeyer B. 2015. *Streptococcus pyogenes* biofilms: formation, biology, and clinical relevance. *Front Cell Infect Microbiol* **5:**15 http://dx.doi.org/10.3389/fcimb.2015.00015.

7. Ferrieri P, Dajani AS, Wannamaker LW, Chapman SS. 1972. Natural history of impetigo. I. Site sequence of acquisition and familial patterns of spread of cutaneous streptococci. *J Clin Invest* **51:**2851–2862 http://dx.doi.org/10.1172/JCI107108.

8. Anthony BF, Kaplan EL, Wannamaker LW, Chapman SS. 1976. The dynamics of streptococcal infections in a defined population of children: serotypes associated with skin and respiratory infections. *Am J Epidemiol* **104:**652–666 http://dx.doi.org/10.1093/oxfordjournals.aje.a112344.

9. Leyden JJ, Stewart R, Kligman AM. 1980. Experimental infections with group A streptococci in humans. *J Invest Dermatol* **75:**196–201 http://dx.doi.org/10.1111/1523-1747.ep12522655.

10. Martin DR, Sriprakash KS. 1996. Epidemiology of group A streptococcal disease in Australia and New Zealand. *Rec Adv Microbiol* **4:**1–40.

11. Potter EV, Svartman M, Mohammed I, Cox R, Poon-King T, Earle DP. 1978. Tropical acute rheumatic fever and associated streptococcal infections compared with concurrent acute glomerulonephritis. *J Pediatr* **92:**325–333 http://dx.doi.org/10.1016/S0022-3476(78)80036-5.

12. Swe PM, Christian LD, Lu HC, Sriprakash KS, Fischer K. 2017. Complement inhibition by Sarcoptes scabiei protects *Streptococcus pyogenes*: an *in vitro* study to unravel the molecular mechanisms behind the poorly understood predilection of *S. pyogenes* to infect mite-induced skin lesions. *PLoS Negl Trop Dis* **11:**e0005437 http://dx.doi.org/10.1371/journal.pntd.0005437.

13. Bowen AC, Tong SY, Chatfield MD, Carapetis JR. 2014. The microbiology of impetigo in indigenous children: associations between *Streptococcus pyogenes*, *Staphylococcus aureus*, scabies, and nasal carriage. *BMC Infect Dis* **14:**727 http://dx.doi.org/10.1186/s12879-014-0727-5.

14. Dajani AS, Ferrieri P, Wannamaker LW. 1972. Natural history of impetigo. II. Etiologic agents and bacterial interactions. *J Clin Invest* **51:**2863–2871 http://dx.doi.org/10.1172/JCI107109.

15. Carapetis JR, Steer AC, Mulholland EK, Weber M. 2005. The global burden of group A streptococcal diseases. *Lancet Infect Dis* **5:**685–694 http://dx.doi.org/10.1016/S1473-3099(05)70267-X.

16. Bowen AC, Mahé A, Hay RJ, Andrews RM, Steer AC, Tong SY, Carapetis JR. 2015. The global epidemiology of impetigo: a systematic review of the population prevalence of impetigo and pyoderma. *PLoS One* **10:**e0136789 http://dx.doi.org/10.1371/journal.pone.0136789.

17. Breese BB, Hall CB. 1978. *Beta Hemolytic Streptococcal Diseases*. Houghton Mifflin, Boston, MA.

18. Martin J. 2016. The *Streptococcus pyogenes* carrier state. *In* Ferretti JJ, Stevens DL, Fischetti VA (ed), Streptococcus pyogenes: *Basic Biology to Clinical Manifestations*. University of Oklahoma Health Sciences Center, Oklahoma City, OK.

19. Nelson GE, Pondo T, Toews KA, Farley MM, Lindegren ML, Lynfield R, Aragon D, Zansky SM, Watt JP, Cieslak PR, Angeles K, Harrison LH, Petit S, Beall B, Van Beneden CA. 2016. Epidemiology of invasive group A streptococcal infections in the United States, 2005-2012. *Clin Infect Dis* **63:**478–486 http://dx.doi.org/10.1093/cid/ciw248.

20. Van Beneden C. 2017. Recent outbreaks of group A streptococcal infection in North America: challenges and new opportunities., The 20th Lancefield International Symposium on Streptococci and Streptococcal Diseases, Denarau Island, Fiji.

21. Fiorentino TR, Beall B, Mshar P, Bessen DE. 1997. A genetic-based evaluation of the principal tissue reservoir for group A streptococci isolated from normally sterile sites. *J Infect Dis* **176:**177–182 http://dx.doi.org/10.1086/514020.

22. Shulman ST, Tanz RR, Kabat W, Kabat K, Cederlund E, Patel D, Li Z, Sakota V, Dale JB, Beall B, US Streptococcal Pharyngitis Surveillance Group. 2004. Group A streptococcal pharyngitis serotype surveillance in North America, 2000-2002. *Clin Infect Dis* **39:**325–332 http://dx.doi.org/10.1086/421949.

23. Tyrrell GJ, Lovgren M, St Jean T, Hoang L, Patrick DM, Horsman G, Van Caeseele P, Sieswerda LE, McGeer A, Laurence RA, Bourgault AM, Low DE. 2010. Epidemic of group A *Streptococcus* M/emm59 causing invasive disease in Canada. *Clin Infect Dis* **51:**1290–1297 http://dx.doi.org/10.1086/657068.

24. Fittipaldi N, Beres SB, Olsen RJ, Kapur V, Shea PR, Watkins ME, Cantu CC, Laucirica DR, Jenkins L, Flores AR, Lovgren M, Ardanuy C, Liñares J, Low DE, Tyrrell GJ, Musser JM. 2012. Full-genome dissection of an epidemic of severe invasive disease caused by a hypervirulent, recently emerged clone of group A *Streptococcus*. *Am J Pathol* **180:**1522–1534 http://dx.doi.org/10.1016/j.ajpath.2011.12.037.

25. Wessels MR, Bronze MS. 1994. Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. *Proc Natl Acad Sci USA* **91:**12238–12242 http://dx.doi.org/10.1073/pnas.91.25.12238.

26. Cywes C, Stamenkovic I, Wessels MR. 2000. CD44 as a receptor for colonization of the pharynx by group A *Streptococcus*. *J Clin Invest* **106:**995–1002 http://dx.doi.org/10.1172/JCI10195.

27. Manetti AG, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F, Gambellini G, Bensi G, Mora M, Edwards AM, Musser JM, Graviss EA, Telford JL, Grandi G, Margarit I. 2007. *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Mol Microbiol* **64:**968–983 http://dx.doi.org/10.1111/j.1365-2958.2007.05704.x.

28. Abbot EL, Smith WD, Siou GP, Chiriboga C, Smith RJ, Wilson JA, Hirst BH, Kehoe MA. 2007. Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. *Cell Microbiol* **9:**1822–1833 http://dx.doi.org/10.1111/j.1462-5822.2007.00918.x.

29. Brouwer S, Barnett TC, Rivera-Hernandez T, Rohde M, Walker MJ. 2016. *Streptococcus pyogenes* adhesion and colonization. *FEBS Lett* **590:**3739–3757 http://dx.doi.org/10.1002/1873-3468.12254.

30. Turner CE, Abbott J, Lamagni T, Holden MT, David S, Jones MD, Game L, Efstratiou A, Sriskandan S. 2015. Emergence of a new highly successful acapsular group A *Streptococcus* clade of genotype emm89 in the United Kingdom. *MBio* **6:**e00622 http://dx.doi.org/10.1128/mBio.00622-15.

31. Zhu L, Olsen RJ, Nasser W, de la Riva Morales I, Musser JM. 2015. Trading capsule for increased cytotoxin production: contribution to virulence of a newly emerged clade of *emm89 Streptococcus pyogenes*. *MBio* **6:**e01378–e15 http://dx.doi.org/10.1128/mBio.01378-15.

32. Bessen DE, Lizano S. 2010. Tissue tropisms in group A streptococcal infections. *Future Microbiol* **5:**623–638 http://dx.doi.org/10.2217/fmb.10.28.

33. Shimomura Y, Okumura K, Murayama SY, Yagi J, Ubukata K, Kirikae T, Miyoshi-Akiyama T. 2011. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics* **12:**17 http://dx.doi.org/10.1186/1471-2164-12-17.

34. Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS. 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals two distinct genetic lineages which do not segregate with their phenotypes or propensity to cause diseases. *Clin Infect Dis* **44:**1142–1154.

35. Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, Gertz RE Jr, Srinivasan V, Lynfield R, Van Beneden C, McGee L, Beall B. 2017. Population and whole genome sequence based characterization of invasive group A streptococci recovered in the United States during 2015. *MBio* **8:**8 http://dx.doi.org/10.1128/mBio.01422-17.

36. Brandt CM, Spellerberg B. 2009. Human infections due to *Streptococcus dysgalactiae* subspecies *equisimilis*. *Clin Infect Dis* **49:**766–772 http://dx.doi.org/10.1086/605085.

37. Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. *Clin Infect Dis* **48:**706–712 http://dx.doi.org/10.1086/597035.

38. Shelburne SA III, Sumby P, Sitkiewicz I, Okorafor N, Granville C, Patel P, Voyich J, Hull R, DeLeo FR, Musser JM. 2006. Maltodextrin utilization plays a key role in the ability of group A *Streptococcus* to colonize the oropharynx. *Infect Immun* **74:**4605–4614 http://dx.doi.org/10.1128/IAI.00477-06.

39. Cook LC, LaSarre B, Federle MJ. 2013. Interspecies communication among commensal and pathogenic streptococci. *MBio* **4:**4 http://dx.doi.org/10.1128/mBio.00382-13.

40. Wescombe PA, Heng NC, Burton JP, Tagg JR. 2010. Something old and something new: an update on the amazing repertoire of bacteriocins produced by *Streptococcus salivarius*. *Probiotics Antimicrob Proteins* **2:**37–45 http://dx.doi.org/10.1007/s12602-009-9026-7.

41. Loughman JA, Caparon M. 2006. Regulation of SpeB in *Streptococcus pyogenes* by pH and NaCl: a model for *in vivo* gene expression. *J Bacteriol* **188:**399–408 http://dx.doi.org/10.1128/JB.188.2.399-408.2006.

42. Henningham A, Döhrmann S, Nizet V, Cole JN. 2015. Mechanisms of group A *Streptococcus* resistance to reactive oxygen species. *FEMS Microbiol Rev* **39:**488–508 http://dx.doi.org/10.1093/femsre/fuu009.

43. Kreikemeyer B, McIver KS, Podbielski A. 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* **11:**224–232 http://dx.doi.org/10.1016/S0966-842X(03)00098-2.

44. Flores AR, Jewell BE, Olsen RJ, Shelburne SA III, Fittipaldi N, Beres SB, Musser JM. 2014. Asymptomatic carriage of group A *Streptococcus* is associated with elimination of capsule production. *Infect Immun* **82:**3958–3967 http://dx.doi.org/10.1128/IAI.01788-14.

45. Flores AR, Olsen RJ, Wunsche A, Kumaraswami M, Shelburne SA III, Carroll RK, Musser JM. 2013. Natural variation in the promoter of the gene encoding the Mga regulator alters host-pathogen interactions in group A *Streptococcus* carrier strains. *Infect Immun* **81:**4128–4138 http://dx.doi.org/10.1128/IAI.00405-13.

46. Anderson R, May R. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, UK.

47. Lancefield RC. 1959. Persistence of type-specific antibodies in man following infection with group A streptococci. *J Exp Med* **110:**271–292 http://dx.doi.org/10.1084/jem.110.2.271.

48. Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AGO, Maggi T, Taddei AR, Grandi G, Telford JL. 2005. Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A* **102:**15641–15646 http://dx.doi.org/10.1073/pnas.0507808102.

49. Bessen D, Fischetti VA. 1990. Synthetic peptide vaccine against mucosal colonization by group A streptococci. I. Protection against a heterologous M serotype with shared C repeat region epitopes. *J Immunol* **145:**1251–1256.

50. Bessen DE, McShan WM, Nguyen SV, Shetty A, Agrawal S, Tettelin H. 2015. Molecular epidemiology and genomics of group A *Streptococcus*. *Infect Genet Evol* **33:**393–418 http://dx.doi.org/10.1016/j.meegid.2014.10.011.

51. Bessen DE. 2016. Tissue tropisms in group A *Streptococcus*: what virulence factors distinguish pharyngitis from impetigo strains? *Curr Opin Infect Dis* **29:**295–303 http://dx.doi.org/10.1097/QCO.0000000000000262.

52. Lancefield RC. 1962. Current knowledge of type-specific M antigens of group A streptococci. *J Immunol* **89:**307–313.

53. Widdowson JP, Maxted WR, Pinney AM. 1971. An M-associated protein antigen (MAP) of group A streptococci. *J Hyg (Lond)* **69:**553–564 http://dx.doi.org/10.1017/S0022172400021823.

54. Widdowson JP, Maxted WR, Pinney AM. 1976. Immunological heterogeneity among the M-associated protein antigens of group-A streptococci. *J Med Microbiol* **9:**73–88 http://dx.doi.org/10.1099/00222615-9-1-73.

55. Bessen D, Jones KF, Fischetti VA. 1989. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *J Exp Med* **169:**269–283 http://dx.doi.org/10.1084/jem.169.1.269.

56. Bessen DE, Fischetti VA. 1990. Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules. *J Exp Med* **172:**1757–1764 http://dx.doi.org/10.1084/jem.172.6.1757.

57. Beall B, Facklam R, Thompson T. 1996. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* **34:**953–958.

58. Beall B. 2017. https://www.cdc.gov/streplab/groupa-strep/index.html. Accessed July 10, 2018.

59. Li Z, Sakota V, Jackson D, Franklin AR, Beall B, Active Bacterial Core Surveillance/Emerging Infections Program Network. 2003. Array of M protein gene subtypes in 1064 recent invasive group A streptococcus isolates recovered from the active bacterial core surveillance. *J Infect Dis* **188:**1587–1592 http://dx.doi.org/10.1086/379050.

60. Hollingshead SK, Readdy TL, Yung DL, Bessen DE. 1993. Structural heterogeneity of the *emm* gene cluster in group A streptococci. *Mol Microbiol* **8:**707–717 http://dx.doi.org/10.1111/j.1365-2958.1993.tb01614.x.

61. Podbielski A. 1993. Three different types of organization of the vir regulon in group A streptococci. *Mol Gen Genet* **237:**287–300.

62. Haanes EJ, Heath DG, Cleary PP. 1992. Architecture of the vir regulons of group A streptococci parallels opacity factor phenotype and M protein class. *J Bacteriol* **174:**4967–4976 http://dx.doi.org/10.1128/jb.174.15.4967-4976.1992.

63. Frost H, Sanderson-Smith M, Walker M, Botteaux A, Smeesters P. 2017. Group A streptococcal M-like proteins: from pathogenesis to vaccine potential. *FEMS Microbiol Rev* **42:**193–204.

64. Hollingshead SK, Arnold J, Readdy TL, Bessen DE. 1994. Molecular evolution of a multigene family in group A streptococci. *Mol Biol Evol* **11:**208–219.

65. Sanderson-Smith M, De Oliveira DM, Guglielmini J, McMillan DJ, Vu T, Holien JK, Henningham A, Steer AC, Bessen DE, Dale JB, Curtis N, Beall BW, Walker MJ, Parker MW, Carapetis JR, Van Melderen L, Sriprakash KS, Smeesters PR, M Protein Study Group. 2014. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. *J Infect Dis* **210:**1325–1338 http://dx.doi.org/10.1093/infdis/jiu260.

66. Baroux N, D'Ortenzio E, Amédéo N, Baker C, Ali Alsuwayyid B, Dupont-Rouzeyrol M, O'Connor O, Steer A, Smeesters PR. 2014. The *emm*-cluster typing system for group A *Streptococcus* identifies epidemiologic similarities across the Pacific region. *Clin Infect Dis* **59:**e84–e92 http://dx.doi.org/10.1093/cid/ciu490.

67. Shulman ST, Tanz RR, Dale JB, Steer AC, Smeesters PR. 2014. Added value of the *emm*-cluster typing system to analyze group A *Streptococcus* epidemiology in high-income settings. *Clin Infect Dis* **59:**1651–1652 http://dx.doi.org/10.1093/cid/ciu649.

68. Seale AC, Davies MR, Anampiu K, Morpeth SC, Nyongesa S, Mwarumba S, Smeesters PR, Efstratiou A, Karugutu R, Mturi N, Williams TN, Scott JA, Kariuki S, Dougan G, Berkley JA. 2016. Invasive group A *Streptococcus* infection among children, rural Kenya. *Emerg Infect Dis* **22:**224–232 http://dx.doi.org/10.3201/eid2202.151358.

69. Smeesters PR, Laho D, Beall B, Steer AC, Van Beneden CA. 2017. Seasonal, geographic, and temporal trends of *emm* clusters associated with invasive group A streptococcal infections in US multistate surveillance. *Clin Infect Dis* **64:**694–695 http://dx.doi.org/10.1093/cid/ciw807.

70. Smeesters PR, McMillan DJ, Sriprakash KS. 2010. The streptococcal M protein: a highly versatile molecule. *Trends Microbiol* **18:**275–282 http://dx.doi.org/10.1016/j.tim.2010.02.007.

71. Maxted WR, Widdowson JP, Fraser CAM, Ball LC, Bassett DCJ. 1973. The use of the serum opacity reaction in the typing of group-A streptococci. *J Med Microbiol* **6:**83–90 http://dx.doi.org/10.1099/00222615-6-1-83.

72. Widdowson JP, Maxted WR, Grant DL, Pinney AM. 1971. The relationship between M-antigen and opacity factor in group A streptococci. *J Gen Microbiol* **65:**69–80 http://dx.doi.org/10.1099/00221287-65-1-69.

73. Pinney AM, Widdowson JP, Maxted WR. 1977. Inhibition of beta-haemolysis by opacity factor in group A streptococci. *J Hyg (Lond)* **78:**355–362 http://dx.doi.org/10.1017/S0022172400056254.

74. Courtney HS, Hasty DL, Li Y, Chiang HC, Thacker JL, Dale JB. 1999. Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 *Streptococcus pyogenes*. *Mol Microbiol* **32:**89–98 http://dx.doi.org/10.1046/j.1365-2958.1999.01328.x.

75. Timmer AM, Kristian SA, Datta V, Jeng A, Gillen CM, Walker MJ, Beall B, Nizet V. 2006. Serum opacity factor promotes group A streptococcal epithelial cell invasion and virulence. *Mol Microbiol* **62:**15–25 http://dx.doi.org/10.1111/j.1365-2958.2006.05337.x.

76. Zhu L, Olsen RJ, Musser JM. 2017. Opacification domain of serum opacity factor inhibits beta-hemolysis and contributes to virulence of *Streptococcus pyogenes*. *MSphere* **2:**e00147-17 http://dx.doi.org/10.1128/mSphereDirect.00147-17.

77. Courtney HS, Pownall HJ. 2010. The structure and function of serum opacity factor: a unique streptococcal virulence determinant that targets high-density lipoproteins. *J Biomed Biotechnol* **2010:**956071 http://dx.doi.org/10.1155/2010/956071.

78. Rakonjac JV, Robbins JC, Fischetti VA. 1995. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. *Infect Immun* **63:**622–631.

79. Jeng A, Sakota V, Li Z, Datta V, Beall B, Nizet V. 2003. Molecular genetic analysis of a group A *Streptococcus* operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. *J Bacteriol* **185:**1208–1217 http://dx.doi.org/10.1128/JB.185.4.1208-1217.2003.

80. Beall B, Gherardi G, Lovgren M, Facklam RR, Forwick BA, Tyrrell GJ. 2000. *emm* and *sof* gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. *Microbiology* **146:**1195–1209 http://dx.doi.org/10.1099/00221287-146-5-1195.

81. Wertz JE, McGregor KF, Bessen DE. 2007. Detecting key structural features within highly recombined genes. *PLOS Comput Biol* **3:**e14 http://dx.doi.org/10.1371/journal.pcbi.0030014.

82. Kratovac Z, Manoharan A, Luo F, Lizano S, Bessen DE. 2007. Population genetics and linkage analysis of loci within the FCT region of *Streptococcus pyogenes*. *J Bacteriol* **189:**1299–1310 http://dx.doi.org/10.1128/JB.01301-06.

83. Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B. 2006. Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and *emm*-gene type with T-protein agglutination pattern and serum opacity factor. *J Med Microbiol* **55:**157–164 http://dx.doi.org/10.1099/jmm.0.46224-0.

84. Griffith F. 1934. The serological classification of *Streptococcus pyogenes*. *J Hyg (Lond)* **34:**542–584 http://dx.doi.org/10.1017/S0022172400043308.

85. Johnson D, Kaplan E, Sramek J, Bicova R, Havlicek J, Havlickova H, Motlova J, Kriz P. 1996. *Laboratory Diagnosis of Group A Streptococcal Infections*. World Health Organization, Geneva, Switzerland.

86. Schneewind O, Jones KF, Fischetti VA. 1990. Sequence and structural characteristics of the trypsin-resistant T6 surface protein of group A streptococci. *J Bacteriol* **172:**3310–3317 http://dx.doi.org/10.1128/jb.172.6.3310-3317.1990.

87. Bessen DE, Kalia A. 2002. Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*. *Infect Immun* **70:**1159–1167 http://dx.doi.org/10.1128/IAI.70.3.1159-1167.2002.

88. Podbielski A, Woischnik M, Leonard BA, Schmidt KH. 1999. Characterization of *nra*, a global negative regulator gene in group A streptococci. *Mol Microbiol* **31:**1051–1064 http://dx.doi.org/10.1046/j.1365-2958.1999.01241.x.

89. Hanski E, Caparon M. 1992. Protein F, a fibronectin-binding protein, is an adhesin of the group A streptococcus *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **89:**6172–6176 http://dx.doi.org/10.1073/pnas.89.13.6172.

90. Efstratiou A. 1983. The serotyping of hospital strains of streptococci belonging to Lancefield group C and G. *J Hyg (Camb)* **90:**71–80.

91. Beres SB, Musser JM. 2007. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. *PLoS One* **2:**e800 http://dx.doi.org/10.1371/journal.pone.0000800.

92. Falugi F, Zingaretti C, Pinto V, Mariani M, Amodeo L, Manetti AG, Capo S, Musser JM, Orefici G, Margarit I, Telford JL, Grandi G, Mora M. 2008. Sequence variation in group A *Streptococcus* pili and association of pilus backbone types with lancefield T serotypes. *J Infect Dis* **198:**1834–1841 http://dx.doi.org/10.1086/593176.

93. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G. 2006. Pili in Gram-positive pathogens. *Nat Rev Microbiol* **4:**509–519 http://dx.doi.org/10.1038/nrmicro1443.

94. Scott JR, Zähner D. 2006. Pili with strong attachments: Gram-positive bacteria do it differently. *Mol Microbiol* **62:**320–330 http://dx.doi.org/10.1111/j.1365-2958.2006.05279.x.

95. Margarit y Ros I. 2016. *Streptococcus pyogenes* pili. *In* Ferretti JJ, Stevens DL, Fischetti VA (ed), Streptococcus pyogenes: *Basic Biology to Clinical Manifestations*. University of Oklahoma Health Sciences Center, Oklahoma City, OK.

96. Steemson JD, Moreland NJ, Williamson D, Morgan J, Carter PE, Proft T. 2014. Survey of the *bp/tee* genes from clinical group A streptococcus isolates in New Zealand: implications for vaccine development. *J Med Microbiol* **63:**1670–1678 http://dx.doi.org/10.1099/jmm.0.080804-0.

97. Lizano S, Luo F, Bessen DE. 2007. Role of streptococcal T antigens in superficial skin infection. *J Bacteriol* **189:**1426–1434 http://dx.doi.org/10.1128/JB.01179-06.

98. Zingaretti C, Falugi F, Nardi-Dei V, Pietrocola G, Mariani M, Liberatori S, Gallotta M, Tontini M, Tani C, Speziale P, Grandi G, Margarit I. 2010. *Streptococcus pyogenes* SpyCEP: a chemokine-inactivating protease with unique structural and biochemical features. *FASEB J* **24:**2839–2848 http://dx.doi.org/10.1096/fj.09-145631.

99. Bao Y, Liang Z, Booyjzsen C, Mayfield JA, Li Y, Lee SW, Ploplis VA, Song H, Castellino FJ. 2014. Unique genomic arrangements in an invasive serotype M23 strain of *Streptococcus pyogenes* identify genes that induce hypervirulence. *J Bacteriol* **196:**4089–4102 http://dx.doi.org/10.1128/JB.02131-14.

100. Luo F, Lizano S, Bessen DE. 2008. Heterogeneity in the polarity of Nra regulatory effects on streptococcal pilus gene transcription and virulence. *Infect Immun* **76:**2490–2497 http://dx.doi.org/10.1128/IAI.01567-07.

101. Bessen DE, Kumar N, Hall GS, Riley DR, Luo F, Lizano S, Ford CN, McShan WM, Nguyen SV, Dunning Hotopp JC, Tettelin H. 2011. Whole-genome association study on tissue tropism phenotypes in group A *Streptococcus*. *J Bacteriol* **193:**6651–6663 http://dx.doi.org/10.1128/JB.05263-11.

102. Loh JMS, Lorenz N, Tsai CJ, Khemlani AHJ, Proft T. 2017. Mucosal vaccination with pili from group A *Streptococcus* expressed on *Lactococcus lactis* generates protective immune responses. *Sci Rep* **7:**7174 http://dx.doi.org/10.1038/s41598-017-07602-0.

103. Musser JM, Hauser AR, Kim MH, Schlievert PM, Nelson K, Selander RK. 1991. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. *Proc Natl Acad Sci U S A* **88:**2668–2672 http://dx.doi.org/10.1073/pnas.88.7.2668.

104. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. *Infect Immun* **69:**2416–2427 http://dx.doi.org/10.1128/IAI.69.4.2416-2427.2001.

105. Jolley K. 2017. *Streptococcus pyogenes* MLST databases. http://pubmlst.org/spyogenes. Accessed Decmber 2017.

106. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* **13:**87 http://dx.doi.org/10.1186/1471-2105-13-87.

107. Bessen DE, McGregor KF, Whatmore AM. 2008. Relationships between *emm* and multilocus sequence types within a global collection of *Streptococcus pyogenes*. *BMC Microbiol* **8:**59 http://dx.doi.org/10.1186/1471-2180-8-59.

108. McGregor KF, Bilek N, Bennett A, Kalia A, Beall B, Carapetis JR, Currie BJ, Sriprakash KS, Spratt BG, Bessen DE. 2004. Group A streptococci from a remote community have novel multilocus genotypes but share *emm* types and housekeeping alleles with isolates from worldwide sources. *J Infect Dis* **189:**717–723 http://dx.doi.org/10.1086/381452.

109. Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S, Lai HS, Lin SP, Qian Y, Jia HG, Najar FZ, Ren Q, Zhu H, Song L, White J, Yuan X, Clifton SW, Roe BA, McLaughlin R. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98:**4658–4663 http://dx.doi.org/10.1073/pnas.071559398.

110. Maruyama F, Watanabe T, Nakagawa I. 2016. *Streptococcus pyogenes* genomics. *In* Ferretti JJ, Stevens DL, Fischetti VA (ed), Streptococcus pyogenes: *Basic Biology to Clinical Manifestations*. University of Oklahoma Health Sciences Center, Oklahoma City, OK.

111. Beres SB, Kachroo P, Nasser W, Olsen RJ, Zhu L, Flores AR, de la Riva I, Paez-Mayorga J, Jimenez FE, Cantu C, Vuopio J, Jalava J, Kristinsson KG, Gottfredsson M, Corander J, Fittipaldi N, Di Luca MC, Petrelli D, Vitali LA, Raiford A, Jenkins L, Musser JM. 2016. Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. *MBio* **7:** e00403-16 http://dx.doi.org/10.1128/mBio.00403-16.

112. Zhu L, Olsen RJ, Nasser W, Beres SB, Vuopio J, Kristinsson KG, Gottfredsson M, Porter AR, DeLeo FR, Musser JM. 2015. A molecular trigger for intercontinental epidemics of group A *Streptococcus*. *J Clin Invest* **125:**3545–3559 http://dx.doi.org/10.1172/JCI82478.

113. Davies MR, Holden MT, Coupland P, Chen JH, Venturini C, Barnett TC, Zakour NL, Tse H, Dougan G, Yuen KY, Walker MJ. 2015. Emergence of scarlet fever *Streptococcus pyogenes emm12* clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nat Genet* **47:**84–87 http://dx.doi.org/10.1038/ng.3147.

114. Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Long SW, Kristinsson KG, Gottfredsson M, Vuopio J, Raisanen K, Caugant DA, Steinbakk M, Low DE, McGeer A, Darenberg J, Henriques-Normark B, Van Beneden CA, Hoffmann S, Musser JM. 2014. Evolutionary pathway to increased virulence and epidemic group A *Streptococcus* disease derived from 3,615 genome sequences. *Proc Natl Acad Sci USA* **111:**E1768–E1776 http://dx.doi.org/10.1073/pnas.1403138111.

115. Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD, Feldgarden M, Birren BW, Fofanov Y, Boos J, Wheaton WD, Honisch C,

Musser JM. 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci USA* **107:**4371–4376 http://dx.doi.org/10.1073/pnas.0911295107.

116. Bao YJ, Liang Z, Mayfield JA, Donahue DL, Carothers KE, Lee SW, Ploplis VA, Castellino FJ. 2016. Genomic characterization of a pattern D *Streptococcus pyogenes emm53* isolate reveals a genetic rationale for invasive skin tropicity. *J Bacteriol* **198:**1712–1724 http://dx.doi.org/10.1128/JB.01019-15.

117. Kapatai G, Coelho J, Platt S, Chalker VJ. 2017. Whole genome sequencing of group A *Streptococcus*: development and evaluation of an automated pipeline for *emm* gene typing. *Peer J* **5:**e3226 http://dx.doi.org/10.7717/peerj.3226.

118. Athey TB, Teatero S, Li A, Marchand-Austin A, Beall BW, Fittipaldi N. 2014. Deriving group A *Streptococcus* typing information from short-read whole-genome sequencing data. *J Clin Microbiol* **52:**1871–1876 http://dx.doi.org/10.1128/JCM.00029-14.

119. Tagini F, Aubert B, Troillet N, Pillonel T, Praz G, Crisinel PA, Prod'hom G, Asner S, Greub G. 2017. Importance of whole genome sequencing for the assessment of outbreaks in diagnostic laboratories: analysis of a case series of invasive *Streptococcus pyogenes* infections. *Eur J Clin Microbiol Infect Dis* **36:**1173–1180 http://dx.doi.org/10.1007/s10096-017-2905-z.

120. Turner CE, Bedford L, Brown NM, Judge K, Török ME, Parkhill J, Peacock SJ. 2017. Community outbreaks of group A *Streptococcus* revealed by genome sequencing. *Sci Rep* **7:**8554 http://dx.doi.org/10.1038/s41598-017-08914-x.

121. Whatmore AM, Kapur V, Sullivan DJ, Musser JM, Kehoe MA. 1994. Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. *Mol Microbiol* **14:**619–631 http://dx.doi.org/10.1111/j.1365-2958.1994.tb01301.x.

122. Wannamaker LW. 1970. Differences between streptococcal infections of the throat and of the skin. I. *N Engl J Med* **282:**23–31 http://dx.doi.org/10.1056/NEJM197001012820106.

123. Bisno AL, Stevens DL. 2010. *Streptococcus pyogenes*. *In* Mandell GL, Douglas RG, Dolin R (ed), *Principles and Practice of Infectious Diseases*, 7th ed. Churchill Livingstone, Philadelphia, PA. http://dx.doi.org/10.1016/B978-0-443-06839-3.00198-3.

124. Bessen DE, Sotir CM, Readdy TL, Hollingshead SK. 1996. Genetic correlates of throat and skin isolates of group A streptococci. *J Infect Dis* **173:**896–900 http://dx.doi.org/10.1093/infdis/173.4.896.

125. Bessen DE, Carapetis JR, Beall B, Katz R, Hibble M, Currie BJ, Collingridge T, Izzo MW, Scaramuzzino DA, Sriprakash KS. 2000. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. *J Infect Dis* **182:**1109–1116 http://dx.doi.org/10.1086/315842.

126. McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE. 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. *J Bacteriol* **186:**4285–4294 http://dx.doi.org/10.1128/JB.186.13.4285-4294.2004.

127. McMillan DJ, Drèze PA, Vu T, Bessen DE, Guglielmini J, Steer AC, Carapetis JR, Van Melderen L, Sriprakash KS, Smeesters PR. 2013. Updated model of group A *Streptococcus* M proteins based on a comprehensive worldwide study. *Clin Microbiol Infect* **19:**E222–E229 http://dx.doi.org/10.1111/1469-0691.12134.

128. Bessen DE. 2009. Population biology of the human restricted pathogen, *Streptococcus pyogenes*. *Infect Genet Evol* **9:**581–593 http://dx.doi.org/10.1016/j.meegid.2009.03.002.

129. Loof TG, Deicke C, Medina E. 2014. The role of coagulation/fibrinolysis during *Streptococcus pyogenes* infection. *Front Cell Infect Microbiol* **4:**128 http://dx.doi.org/10.3389/fcimb.2014.00128.

130. Rogers S, Commons R, Danchin MH, Selvaraj G, Kelpie L, Curtis N, Robins-Browne R, Carapetis JR. 2007. Strain prevalence, rather than innate virulence potential, is the major factor responsible for an increase in serious group A streptococcus infections. *J Infect Dis* **195:**1625–1633 http://dx.doi.org/10.1086/513875.

131. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. 2009. Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infect Dis* **9:**611–616 http://dx.doi.org/10.1016/S1473-3099(09)70178-1.

132. Efstratiou A, Lamagni T. 2016. Epidemiology of *Streptococcus pyogenes*. *In* Ferretti JJ, Stevens DL, Fischetti VA (ed), Streptococcus pyogenes: *Basic Biology to Clinical Manifestations*. University of Oklahoma Health Sciences Center, Oklahoma City, OK.

133. van Beneden C. 2017. Active bacterial core surveillance (ABCs). https://www.cdc.gov/abcs/reports-findings/surv-reports.html. Accessed December 2017.

134. McGregor KF, Spratt BG. 2005. Identity and prevalence of multilocus sequence typing-defined clones of group A streptococci within a hospital setting. *J Clin Microbiol* **43:**1963–1967 http://dx.doi.org/10.1128/JCM.43.4.1963-1967.2005.

135. Engelthaler DM, Valentine M, Bowers J, Pistole J, Driebe EM, Terriquez J, Nienstadt L, Carroll M, Schumacher M, Ormsby ME, Brady S, Livar E, Yazzie D, Waddell V, Peoples M, Komatsu K, Keim P. 2016. Hypervirulent *emm59* clone in invasive group A *Streptococcus* outbreak, southwestern United States. *Emerg Infect Dis* **22:**734–738 http://dx.doi.org/10.3201/eid2204.151582.

136. Wasserzug O, Valinsky L, Klement E, Bar-Zeev Y, Davidovitch N, Orr N, Korenman Z, Kayouf R, Sela T, Ambar R, Derazne E, Dagan R, Zarka S. 2009. A cluster of ecthyma outbreaks caused by a single clone of invasive and highly infective *Streptococcus pyogenes*. *Clin Infect Dis* **48:**1213–1219 http://dx.doi.org/10.1086/597770.

137. Kwiatkowska RM, Manley P, Sims B, Lamagni T, Ready D, Coelho J, Alsaffar L, Beck CR, Neely F, Outbreak Control Team. 2017. Outbreak of group A Streptococcus emm94.0 affecting people who inject drugs in southwest England, April 2017. *Am J Infect Control* **46:**238–240.

138. Bundle N, Bubba L, Coelho J, Kwiatkowska R, Cloke R, King S, Rajan-Iyer J, Courtney-Pillinger M, Beck CR, Hope V, Lamagni T, Brown CS, Jermacane D, Glass R, Desai M, Gobin M, Balasegaram S, Anderson C. 2017. Ongoing outbreak of invasive and non-invasive disease due to group A *Streptococcus* (GAS) type *emm66* among homeless and people who inject drugs in England and Wales, January to December 2016. *Euro Surveill* **22:**30446 http://dx.doi.org/10.2807/1560-7917.ES.2017.22.3.30446.

139. Veasy LG, Tani LY. 2005. A new look at acute rheumatic mitral regurgitation. *Cardiol Young* **15:**568–577 http://dx.doi.org/10.1017/S104795110500171X.

140. Bisno AL. 2010. Nonsuppurative poststreptococcal sequelae: rheumatic fever and glomerulonephritis. *In* Mandell GL, Bennett JE, Dolin R (ed), *Principles and Practice of Infectious Diseases*, 7th ed. Churchill Livingstone, Philadelphia, PA. http://dx.doi.org/10.1016/B978-0-443-06839-3.00199-5

141. Zühlke LJ, Beaton A, Engel ME, Hugo-Hamman CT, Karthikeyan G, Katzenellenbogen JM, Ntusi N, Ralph AP, Saxena A, Smeesters PR, Watkins D, Zilla P, Carapetis J. 2017. Group A *Streptococcus*, acute rheumatic fever and rheumatic heart disease: epidemiology and clinical considerations. *Curr Treat Options Cardiovasc Med* **19:**15 http://dx.doi.org/10.1007/s11936-017-0513-y.

142. Martin WJ, Steer AC, Smeesters PR, Keeble J, Inouye M, Carapetis J, Wicks IP. 2015. Post-infectious group A streptococcal autoimmune syndromes and the heart. *Autoimmun Rev* **14:**710–725 http://dx.doi.org/10.1016/j.autrev.2015.04.005.

143. Parks T, Mirabel MM, Kado J, Auckland K, Nowak J, Rautanen A, Mentzer AJ, Marijon E, Jouven X, Perman ML, Cua T, Kauwe JK, Allen JB, Taylor H, Robson KJ, Deane CM, Steer AC, Hill AVS, Allen L, Allen M, Braunstein C, Colquhoun SM, Jewine A, Ah Kee M, Kumar R, John Martin W, Mataika R, Nadra M, Nadu S, Naseri T, Noël B, Simon N, Ward B, Pacific Islands Rheumatic Heart Disease Genetics Network. 2017. Association between a common immunoglobulin heavy chain allele

and rheumatic heart disease risk in Oceania. *Nat Commun* **8:**14946 http://dx.doi.org/10.1038/ncomms14946.

144. Carapetis JR. 2007. Rheumatic heart disease in developing countries. *N Engl J Med* **357:**439–441 http://dx.doi.org/10.1056/NEJMp078039.

145. Agarwal BL. 1981. Rheumatic heart disease unabated in developing countries. *Lancet* **2:**910–911 http://dx.doi.org/10.1016/S0140-6736(81)91400-8.

146. Watkins DA, Johnson CO, Colquhoun SM, Karthikeyan G, Beaton A, Bukhman G, Forouzanfar MH, Longenecker CT, Mayosi BM, Mensah GA, Nascimento BR, Ribeiro ALP, Sable CA, Steer AC, Naghavi M, Mokdad AH, Murray CJL, Vos T, Carapetis JR, Roth GA. 2017. Global, regional, and national burden of rheumatic heart disease, 1990-2015. *N Engl J Med* **377:**713–722 http://dx.doi.org/10.1056/NEJMoa1603693.

147. Martins TB, Veasy LG, Hill HR. 2006. Antibody responses to group A streptococcal infections in acute rheumatic fever. *Pediatr Infect Dis J* **25:**832–837 http://dx.doi.org/10.1097/01.inf.0000232634.55883.d0.

148. Soler-Soler J, Galve E. 2000. Worldwide perspective of valve disease. *Heart* **83:**721–725 http://dx.doi.org/10.1136/heart.83.6.721.

149. Veasy LG, Wiedmeier SE, Orsmond GS, Ruttenberg HD, Boucek MM, Roth SJ, Tait VF, Thompson JA, Daly JA, Kaplan EL, Hill HR. 1987. Resurgence of acute rheumatic fever in the intermountain area of the United States. *N Engl J Med* **316:**421–427 http://dx.doi.org/10.1056/NEJM198702193160801.

150. Stollerman GH. 1991. Rheumatogenic streptococci and autoimmunity. *Clin Immunol Immunopathol* **61:**131–142 http://dx.doi.org/10.1016/S0090-1229(05)80019-4.

151. Bisno AL, Pearce IA, Stollerman GH. 1977. Streptococcal infections that fail to cause recurrences of rheumatic fever. *J Infect Dis* **136:**278–285 http://dx.doi.org/10.1093/infdis/136.2.278.

152. Stollerman GH. 1971. Rheumatogenic and nephritogenic streptococci. *Circulation* **43:**915–921 http://dx.doi.org/10.1161/01.CIR.43.6.915.

153. Bisno AL. 1980. The concept of rheumatogenic and nonrheumatogenic group A streptococci, p 789–803. *In* Read SE, Zabriskie JB (ed), *Streptococcal Diseases and the Immune Response*. Academic Press, New York, NY.

154. Johnson DR, Stevens DL, Kaplan EL. 1992. Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever, or uncomplicated pharyngitis. *J Infect Dis* **166:**374–382 http://dx.doi.org/10.1093/infdis/166.2.374.

155. Kuttner AG, Krumwiede E. 1941. Observations on the effect of streptococcal upper respiratory infections on rheumatic children: a three year study. *J Clin Invest* **20:**273–287 http://dx.doi.org/10.1172/JCI101220.

156. Bisno AL, Pearce IA, Wall HP, Moody MD, Stollerman GH. 1970. Contrasting epidemiology of acute rheumatic fever and acute glomerulonephritis. *N Engl J Med* **283:**561–565 http://dx.doi.org/10.1056/NEJM197009102831103.

157. Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. 2006. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. *Clin Infect Dis* **42:**441–447 http://dx.doi.org/10.1086/499812.

158. Erdem G, Mizumoto C, Esaki D, Reddy V, Kurahara D, Yamaga K, Abe L, Johnson D, Yamamoto K, Kaplan EL. 2007. Group A streptococcal isolates temporally associated with acute rheumatic fever in Hawaii: differences from the continental United States. *Clin Infect Dis* **45:**e20–e24 http://dx.doi.org/10.1086/519384.

159. McDonald M, Currie BJ, Carapetis JR. 2004. Acute rheumatic fever: a chink in the chain that links the heart to the throat? *Lancet Infect Dis* **4:**240–245 http://dx.doi.org/10.1016/S1473-3099(04)00975-2.

160. Parks T, Smeesters PR, Steer AC. 2012. Streptococcal skin infection and rheumatic heart disease. *Curr Opin Infect Dis* **25:**145–153 http://dx.doi.org/10.1097/QCO.0b013e3283511d27.

161. Williamson DA, Smeesters PR, Steer AC, Steemson JD, Ng AC, Proft T, Fraser JD, Baker MG, Morgan J, Carter PE, Moreland NJ. 2015. M-protein analysis of *Streptococcus pyogenes* isolates associated with acute rheumatic fever in New Zealand. *J Clin Microbiol* **53:**3618–3620 http://dx.doi.org/10.1128/JCM.02129-15.

162. O'Sullivan L, Moreland NJ, Webb RH, Upton A, Wilson NJ. 2017. Acute rheumatic fever after group A *Streptococcus pyoderma* and group G *Streptococcus pharyngitis*. *Pediatr Infect Dis J* **36:**692–694 http://dx.doi.org/10.1097/INF.0000000000001558.

163. Hollingshead SK, Fischetti VA, Scott JR. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* **207:**196–203 http://dx.doi.org/10.1007/BF00331578.

164. Lannergård J, Gustafsson MC, Waldemarsson J, Norrby-Teglund A, Stålhammar-Carlemalm M, Lindahl G. 2011. The hypervariable region of *Streptococcus pyogenes* M protein escapes antibody attack by antigenic variation and weak immunogenicity. *Cell Host Microbe* **10:**147–157 http://dx.doi.org/10.1016/j.chom.2011.06.011.

165. Jaggi P, Tanz RR, Beall B, Shulman ST. 2005. Age influences the *emm* type distribution of pediatric group A streptococcal pharyngeal isolates. *Pediatr Infect Dis J* **24:**1089–1092 http://dx.doi.org/10.1097/01.inf.0000190023.89759.96.

166. Stewart CM, Buffalo CZ, Valderrama JA, Henningham A, Cole JN, Nizet V, Ghosh P. 2016. Coiled-coil destabilizing residues in the group A *Streptococcus* M1 protein are required for functional interaction. *Proc Natl Acad Sci U S A* **113:**9515–9520 http://dx.doi.org/10.1073/pnas.1606160113.

167. Persson J, Beall B, Linse S, Lindahl G. 2006. Extreme sequence divergence but conserved ligand-binding specificity in *Streptococcus pyogenes* M protein. *PLoS Pathog* **2:**e47 http://dx.doi.org/10.1371/journal.ppat.0020047.

168. Buffalo CZ, Bahn-Suh AJ, Hirakis SP, Biswas T, Amaro RE, Nizet V, Ghosh P. 2016. Conserved patterns hidden within group A *Streptococcus* M protein hypervariability recognize human C4b-binding protein. *Nat Microbiol* **1:**16155 http://dx.doi.org/10.1038/nmicrobiol.2016.155.

169. Rouchon CN, Ly AT, Noto JP, Luo F, Lizano S, Bessen DE. 2017. Incremental contributions of FbaA and other impetigo-associated surface proteins to fitness and virulence of a classical group A streptococcal skin strain. *Infect Immun* **85:**e00374–e17 http://dx.doi.org/10.1128/IAI.00374-17.

170. Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, Moreland NJ, Mulholland EK, Schodel F, Smeesters PR. 2016. Status of research and development of vaccines for *Streptococcus pyogenes*. *Vaccine* **34:**2953–2958 http://dx.doi.org/10.1016/j.vaccine.2016.03.073.

171. Moreland NJ, Waddington CS, Williamson DA, Sriskandan S, Smeesters PR, Proft T, Steer AC, Walker MJ, Baker EN, Baker MG, Lennon D, Dunbar R, Carapetis J, Fraser JD. 2014. Working towards a group A streptococcal vaccine: report of a collaborative Trans-Tasman workshop. *Vaccine* **32:**3713–3720 http://dx.doi.org/10.1016/j.vaccine.2014.05.017.

172. Schödel F, Moreland NJ, Wittes JT, Mulholland K, Frazer I, Steer AC, Fraser JD, Carapetis J. 2017. Clinical development strategy for a candidate group A streptococcal vaccine. *Vaccine* **35:**2007–2014 http://dx.doi.org/10.1016/j.vaccine.2017.02.060.

173. Steer AC, Dale JB, Carapetis JR. 2013. Progress toward a global group a streptococcal vaccine. *Pediatr Infect Dis J* **32:**180–182 http://dx.doi.org/10.1097/INF.0b013e318281da11.

174. Dale JB, Fischetti VA, Carapetis JR, Steer AC, Sow S, Kumar R, Mayosi BM, Rubin FA, Mulholland K, Hombach JM, Schödel F, Henao-Restrepo AM. 2013. Group A streptococcal vaccines: paving a path for accelerated development. *Vaccine* **31**(Suppl 2)**:**B216–B222 http://dx.doi.org/10.1016/j.vaccine.2012.09.045.

175. Pandey M, Batzloff MR, Good MF. 2012. Vaccination against rheumatic heart disease: a review of current research strategies and challenges. *Curr Infect Dis Rep* **14:**381–390 http://dx.doi.org/10.1007/s11908-012-0263-7.

176. Good MF, Batzloff MR, Pandey M. 2013. Strategies in the development of vaccines to prevent infections with group A streptococcus. *Hum Vaccin Immunother* **9:**2393–2397 http://dx.doi.org/10.4161/hv.25506.

177. Henningham A, Gillen CM, Walker MJ. 2013. Group a streptococcal vaccine candidates: potential for the development of a human vaccine. *Curr Top Microbiol Immunol* **368:**207–242 http://dx.doi.org/10.1007/82_2012_284.

178. Henningham A, Ericsson DJ, Langer K, Casey LW, Jovcevski B, Chhatwal GS, Aquilina JA, Batzloff MR, Kobe B, Walker MJ. 2013. Structure-informed design of an enzymatically inactive vaccine component for group A *Streptococcus*. *MBio* **4:**e00509-13 http://dx.doi.org/10.1128/mBio.00509-13.

179. McNeil SA, Halperin SA, Langley JM, Smith B, Warren A, Sharratt GP, Baxendale DM, Reddish MA, Hu MC, Stroop SD, Linden J, Fries LF, Vink PE, Dale JB. 2005. Safety and immunogenicity of 26-valent group A streptococcus vaccine in healthy adult volunteers. *Clin Infect Dis* **41:**1114–1122 http://dx.doi.org/10.1086/444458.

180. Pandey M, Wykes MN, Hartas J, Good MF, Batzloff MR. 2013. Long-term antibody memory induced by synthetic peptide vaccination is protective against *Streptococcus pyogenes* infection and is independent of memory T cell help. *J Immunol* **190:**2692–2701 http://dx.doi.org/10.4049/jimmunol.1202333.

181. Bauer MJ, Georgousakis MM, Vu T, Henningham A, Hofmann A, Rettel M, Hafner LM, Sriprakash KS, McMillan DJ. 2012. Evaluation of novel *Streptococcus pyogenes* vaccine candidates incorporating multiple conserved sequences from the C-repeat region of the M-protein. *Vaccine* **30:**2197–2205 http://dx.doi.org/10.1016/j.vaccine.2011.12.115.

182. McNeilly C, Cosh S, Vu T, Nichols J, Henningham A, Hofmann A, Fane A, Smeesters PR, Rush CM, Hafner LM, Ketheesan N, Sriprakash KS, McMillan DJ. 2016. Predicted coverage and immuno-safety of a recombinant C-repeat region based *Streptococcus pyogenes* vaccine candidate. *PLoS One* **11:**e0156639 http://dx.doi.org/10.1371/journal.pone.0156639.

183. Bessen D, Fischetti VA. 1988. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. *Infect Immun* **56:**2666–2672.

184. De Amicis KM, Freschi de Barros S, Alencar RE, Postol E, Martins CD, Arcuri HA, Goulart C, Kalil J, Guilherme L. 2013. Analysis of the coverage capacity of the StreptInCor candidate vaccine against *Streptococcus pyogenes*. *Vaccine* **32:**4104–4110.

185. Fischetti VA, Hodges WM, Hruby DE. 1989. Protection against streptococcal pharyngeal colonization with a vaccinia: M protein recombinant. *Science* **244:**1487–1490 http://dx.doi.org/10.1126/science.2660266.

186. Dale JB, Penfound TA, Chiang EY, Walton WJ. 2011. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine* **29:**8175–8178 http://dx.doi.org/10.1016/j.vaccine.2011.09.005.

187. Denny FW Jr, Perry WD, Wannamaker LW. 1957. Type-specific streptococcal antibody. *J Clin Invest* **36:**1092–1100 http://dx.doi.org/10.1172/JCI103504.

188. Wannamaker LW, Denny FW, Perry WD, Siegel AC, Rammelkamp CH Jr. 1953. Studies on immunity to streptococcal infections in man. *AMA Am J Dis Child* **86:**347–348.

189. Watson RF, Rothbard S, Swift HF. 1946. Type-specific protection and immunity following intranasal inoculation of monkeys with group A hemolytic streptococci. *J Exp Med* **84:**127–142 http://dx.doi.org/10.1084/jem.84.2.127.

190. Lancefield RC. 1959. Persistence of type-specific antibodies in man following infection with group A streptococci. *J Exp Med* **110:**271–292 http://dx.doi.org/10.1084/jem.110.2.271.

191. Smeesters PR, Vergison A, Campos D, de Aguiar E, Miendje Deyi VY, Van Melderen L. 2006. Differences between Belgian and Brazilian group A *Streptococcus* epidemiologic landscape. *PLoS One* **1:**e10 http://dx.doi.org/10.1371/journal.pone.0000010.

192. Smeesters PR, McMillan DJ, Sriprakash KS, Georgousakis MM. 2009. Differences among group A streptococcus epidemiological landscapes: consequences for M protein-based vaccines? *Expert Rev Vaccines* **8:**1705–1720 http://dx.doi.org/10.1586/erv.09.133.

193. Tewodros W, Kronvall G. 2005. M protein gene (*emm* type) analysis of group A beta-hemolytic streptococci from Ethiopia reveals unique patterns. *J Clin Microbiol* **43:**4369–4376 http://dx.doi.org/10.1128/JCM.43.9.4369-4376.2005.

194. Dey N, McMillan DJ, Yarwood PJ, Joshi RM, Kumar R, Good MF, Sriprakash KS, Vohra H. 2005. High diversity of group A streptococcal *emm* types in an Indian community: the need to tailor multivalent vaccines. *Clin Infect Dis* **40:**46–51 http://dx.doi.org/10.1086/426443.

195. Steer AC, Magor G, Jenney AW, Kado J, Good MF, McMillan D, Batzloff M, Carapetis JR. 2009. *emm* and C-repeat region molecular typing of beta-hemolytic streptococci in a tropical country: implications for vaccine development. *J Clin Microbiol* **47:**2502–2509 http://dx.doi.org/10.1128/JCM.00312-09.

196. Smeesters PR, Mardulyn P, Vergison A, Leplae R, Van Melderen L. 2008. Genetic diversity of group A *Streptococcus* M protein: implications for typing and vaccine development. *Vaccine* **26:**5835–5842 http://dx.doi.org/10.1016/j.vaccine.2008.08.037.

197. Smeesters PR, Dramaix M, Van Melderen L. 2010. The *emm*-type diversity does not always reflect the M protein genetic diversity: is there a case for designer vaccine against GAS. *Vaccine* **28:**883–885 http://dx.doi.org/10.1016/j.vaccine.2009.10.137.

198. Dale JB, Smeesters PR, Courtney HS, Penfound TA, Hohn CM, Smith JC, Baudry JY. 2017. Structure-based design of broadly protective group a streptococcal M protein-based vaccines. *Vaccine* **35:**19–26 http://dx.doi.org/10.1016/j.vaccine.2016.11.065.

199. Frost HR, Laho D, Sanderson-Smith ML, Licciardi P, Donath S, Curtis N, Kado J, Dale JB, Steer AC, Smeesters PR. 2017. Immune cross-opsonization within *emm* clusters following group A *Streptococcus* skin infection: broadening the scope of type-specific immunity. *Clin Infect Dis* **65:**1523–1531 http://dx.doi.org/10.1093/cid/cix599.

200. Tsoi SK, Smeesters PR, Frost HR, Licciardi P, Steer AC. 2015. Correlates of protection for M protein-based vaccines against group A *Streptococcus*. *J Immunol Res* **2015:**167089 http://dx.doi.org/10.1155/2015/167089.

201. Hausdorff WP, Van Dyke MK, Van Effelterre T. 2012. Serotype replacement after pneumococcal vaccination. *Lancet* **379:**1387–1388, author reply 1388–1389 http://dx.doi.org/10.1016/S0140-6736(12)60589-3.

202. Lipsitch M. 1997. Vaccination against colonizing bacteria with multiple serotypes. *Proc Natl Acad Sci USA* **94:**6571–6576 http://dx.doi.org/10.1073/pnas.94.12.6571.

203. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186:**1518–1530 http://dx.doi.org/10.1128/JB.186.5.1518-1530.2004.

204. Friães A, Machado MP, Pato C, Carriço J, Melo-Cristino J, Ramirez M. 2015. Emergence of the same successful clade among distinct populations of *emm89 Streptococcus pyogenes* in multiple geographic regions. *MBio* **6:**e01780–e15 http://dx.doi.org/10.1128/mBio.01780-15.

205. Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* **7:**30 http://dx.doi.org/10.1186/1471-2180-7-30.

206. Feil EJ, Holmes EC, Bessen DE, Chan M-S, Day NPJ, Enright MC, Goldstein R, Hood DW, Kalia A, Moore CE, Zhou J, Spratt BG. 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* **98:**182–187 http://dx.doi.org/10.1073/pnas.98.1.182.

207. Hanage WP, Fraser C, Spratt BG. 2006. The impact of homologous recombination on the generation of diversity in bacteria. *J Theor Biol* **239:**210–219 http://dx.doi.org/10.1016/j.jtbi.2005.08.035.

208. Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3:**199–208 http://dx.doi.org/10.1038/ismej.2008.93.

209. Carapetis J, Gardiner D, Currie B, Mathews JD. 1995. Multiple strains of *Streptococcus pyogenes* in skin sores of aboriginal Australians. *J Clin Microbiol* **33:**1471–1472.

210. Roberts AL, Connolly KL, Kirse DJ, Evans AK, Poehling KA, Peters TR, Reid SD. 2012. Detection of group A *Streptococcus* in tonsils from pediatric patients reveals high rate of asymptomatic streptococcal carriage. *BMC Pediatr* **12:**3 http://dx.doi.org/10.1186/1471-2431-12-3.

211. Akiyama H, Morizane S, Yamasaki O, Oono T, Iwatsuki K. 2003. Assessment of *Streptococcus pyogenes* microcolony formation in infected skin by confocal laser scanning microscopy. *J Dermatol Sci* **32:**193–199 http://dx.doi.org/10.1016/S0923-1811(03)00096-3.

212. Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A. 2012. Replacing and additive horizontal gene transfer in *Streptococcus*. *Mol Biol Evol* **29:**3309–3320 http://dx.doi.org/10.1093/molbev/mss138.

213. Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS. 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals 2 distinct genetic lineages that do not segregate with their phenotypes or propensity to cause diseases. *Clin Infect Dis* **44:**1442–1454 http://dx.doi.org/10.1086/516780.

214. McMillan DJ, Bessen DE, Pinho M, Ford C, Hall GS, Melo-Cristino J, Ramirez M. 2010. Population genetics of *Streptococcus dysgalactiae* subspecies *equisimilis* reveals widely dispersed clones and extensive recombination.*PLoS One* **5:**e11741 http://dx.doi.org/10.1371/journal.pone.0011741.

215. Ahmad Y, Gertz RE Jr, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. *J Clin Microbiol* **47:**2046–2054 http://dx.doi.org/10.1128/JCM.00246-09.

216. Towers RJ, Gal D, McMillan D, Sriprakash KS, Currie BJ, Walker MJ, Chhatwal GS, Fagan PK. 2004. Fibronectin-binding protein gene recombination and horizontal transfer between group A and G streptococci. *J Clin Microbiol* **42:**5357–5361 http://dx.doi.org/10.1128/JCM.42.11.5357-5361.2004.

217. Bessen DE, Manoharan A, Luo F, Wertz JE, Robinson DA. 2005. Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen *Streptococcus pyogenes*. *J Bacteriol* **187:**4163–4172 http://dx.doi.org/10.1128/JB.187.12.4163-4172.2005.

218. Kalia A, Bessen DE. 2004. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. *J Bacteriol* **186:**110–121 http://dx.doi.org/10.1128/JB.186.1.110-121.2004.

219. Giovanetti E, Brenciani A, Morroni G, Tiberi E, Pasquaroli S, Mingoia M, Varaldo PE. 2014. Transduction of the *Streptococcus pyogenes* bacteriophage Φm46.1, carrying resistance genes *mef*(A) and *tet*(O), to other *Streptococcus* species. *Front Microbiol* **5:**746.

220. Mashburn-Warren L, Morrison DA, Federle MJ. 2012. The cryptic competence pathway in *Streptococcus pyogenes* is controlled by a peptide pheromone. *J Bacteriol* **194:**4589–4600 http://dx.doi.org/10.1128/JB.00830-12.

221. Woodbury RL, Wang X, Moran CP Jr. 2006. Sigma X induces competence gene expression in *Streptococcus pyogenes*. *Res Microbiol* **157:**851–856 http://dx.doi.org/10.1016/j.resmic.2006.07.002.

222. Marks LR, Mashburn-Warren L, Federle MJ, Hakansson AP. 2014. *Streptococcus pyogenes* biofilm growth *in vitro* and *in vivo* and its role in colonization, virulence, and genetic exchange. *J Infect Dis* **210:**25–34 http://dx.doi.org/10.1093/infdis/jiu058.

223. Scott J, Thompson-Mayberry P, Lahmamsi S, King CJ, McShan WM. 2008. Phage-associated mutator phenotype in group A streptococcus. *J Bacteriol* **190:**6290–6301 http://dx.doi.org/10.1128/JB.01569-07.

224. Kalia A, Spratt BG, Enright MC, Bessen DE. 2002. Influence of recombination and niche separation on the population genetic structure of the pathogen *Streptococcus pyogenes*. *Infect Immun* **70:**1971–1983 http://dx.doi.org/10.1128/IAI.70.4.1971-1983.2002.

225. Gupta S, Maiden MC, Feavers IM, Nee S, May RM, Anderson RM. 1996. The maintenance of strain structure in populations of recombining infectious agents. *Nat Med* **2:**437–442 http://dx.doi.org/10.1038/nm0496-437.

226. Bessen DE, Hollingshead SK. 1994. Allelic polymorphism of *emm* loci provides evidence for horizontal gene spread in group A streptococci. *Proc Natl Acad Sci USA* **91:**3280–3284 http://dx.doi.org/10.1073/pnas.91.8.3280.

227. Bessen DE, Izzo MW, McCabe EJ, Sotir CM. 1997. Two-domain motif for IgG-binding activity by group A streptococcal *emm* gene products. *Gene* **196:**75–82 http://dx.doi.org/10.1016/S0378-1119(97)00201-1.

228. Bessen DE. 1994. Localization of immunoglobulin A-binding sites within M or M-like proteins of group A streptococci. *Infect Immun* **62:**1968–1974.

229. O'Toole PW, Stenberg L, Rissler M, Lindahl G. 1992. Two major classes in the M protein family in group A streptococci. *Proc Natl Acad Sci USA* **89:**8661–8665 http://dx.doi.org/10.1073/pnas.89.18.8661.

230. Johnsson E, Andersson G, Lindahl G, Hedén L-O. 1994. Identification of the IgA-binding region in streptococcal protein Arp. *J Immunol* **153:**3557–3564.

231. Bessen DE, McGregor KF, Whatmore AM. 2008. Relationships between *emm* and multilocus sequence types within a global collection of *Streptococcus pyogenes*. *BMC Microbiol* **8:**59 http://dx.doi.org/10.1186/1471-2180-8-59.

232. Kalia A, Spratt BG, Enright MC, Bessen DE. 2002. Influence of recombination and niche separation on the population genetic structure of the pathogen *Streptococcus pyogenes*. *Infect Immun* **70:**1971–1983 http://dx.doi.org/10.1128/IAI.70.4.1971-1983.2002.

233. Beres SB, Sylva GL, Sturdevant DE, Granville CN, Liu M, Ricklefs SM, Whitney AR, Parkins LD, Hoe NP, Adams GJ, Low DE, DeLeo FR, McGeer A, Musser JM. 2004. Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. *Proc Natl Acad Sci U S A* **101:**11833–11838 http://dx.doi.org/10.1073/pnas.0404163101.

234. Shea PR, Beres SB, Flores AR, Ewbank AL, Gonzalez-Lugo JH, Martagon-Rosado AJ, Martinez-Gutierrez JC, Rehman HA, Serrano-Gonzalez M, Fittipaldi N, Ayers SD, Webb P, Willey BM, Low DE, Musser JM. 2011. Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proc Natl Acad Sci U S A* **108:**5039–5044 http://dx.doi.org/10.1073/pnas.1016282108.

235. Olsen RJ, Laucirica DR, Watkins ME, Feske ML, Garcia-Bustillos JR, Vu C, Cantu C, Shelburne SA III, Fittipaldi N, Kumaraswami M, Shea PR, Flores AR, Beres SB, Lovgren M, Tyrrell GJ, Efstratiou A, Low DE, Van Beneden CA, Musser JM. 2012. Polymorphisms in regulator of protease B (RopB) alter disease phenotype and strain virulence of serotype M3 group A *Streptococcus*. *J Infect Dis* **205:**1719–1729 http://dx.doi.org/10.1093/infdis/jir825.

236. Ikebe T, Ato M, Matsumura T, Hasegawa H, Sata T, Kobayashi K, Watanabe H. 2010. Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. *PLoS Pathog* **6:**e1000832 http://dx.doi.org/10.1371/journal.ppat.1000832.

237. Riddle DJ, Bessen DE, Caparon MG. 2010. Variation in *Streptococcus pyogenes* NAD+ glycohydrolase is associated with tissue tropism. *J Bacteriol* **192:**3735–3746 http://dx.doi.org/10.1128/JB.00234-10.

238. Sanson M, O'Neill BE, Kachroo P, Anderson JR, Flores AR, Valson C, Cantu CC, Makthal N, Karmonik C, Fittipaldi N, Kumaraswami M, Musser JM, Olsen RJ. 2015. A naturally occurring single amino acid replacement in multiple gene regulator of group A *Streptococcus* significantly increases virulence. *Am J Pathol* **185:**462–471 http://dx.doi.org/10.1016/j.ajpath.2014.10.018.

239. Walker MJ, Hollands A, Sanderson-Smith ML, Cole JN, Kirk JK, Henningham A, McArthur JD, Dinkla K, Aziz RK, Kansal RG, Simpson AJ, Buchanan JT, Chhatwal GS, Kotb M, Nizet V. 2007. DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nat Med* **13:**981–985 http://dx.doi.org/10.1038/nm1612.

240. Flores AR, Jewell BE, Yelamanchili D, Olsen RJ, Musser JM. 2015. A single amino acid replacement in the sensor kinase LiaS contributes to a carrier phenotype in group A *Streptococcus*. *Infect Immun* **83:**4237–4246 http://dx.doi.org/10.1128/IAI.00656-15.

241. Lau EH, Nishiura H, Cowling BJ, Ip DK, Wu JT. 2012. Scarlet fever outbreak, Hong Kong, 2011. *Emerg Infect Dis* **18:**1700–1702 http://dx.doi.org/10.3201/eid1810.120062.

242. Luk EY, Lo JY, Li AZ, Lau MC, Cheung TK, Wong AY, Wong MM, Wong CW, Chuang SK, Tsang T. 2012. Scarlet fever epidemic, Hong Kong, 2011. *Emerg Infect Dis* **18:**1658–1661 http://dx.doi.org/10.3201/eid1810.111900.

243. Hsieh YC, Huang YC. 2011. Scarlet fever outbreak in Hong Kong, 2011. *J Microbiol Immunol Infect* **44:**409–411 http://dx.doi.org/10.1016/j.jmii.2011.07.003.

244. Tse H, Bao JY, Davies MR, Maamary P, Tsoi HW, Tong AH, Ho TC, Lin CH, Gillen CM, Barnett TC, Chen JH, Lee M, Yam WC, Wong CK, Ong CL, Chan YW, Wu CW, Ng T, Lim WW, Tsang TH, Tse CW, Dougan G, Walker MJ, Lok S, Yuen KY. 2012. Molecular characterization of the 2011 Hong Kong scarlet fever outbreak. *J Infect Dis* **206:**341–351 http://dx.doi.org/10.1093/infdis/jis362.

245. Lamagni T, Guy R, Chand M, Henderson KL, Chalker V, Lewis J, Saliba V, Elliot AJ, Smith GE, Rushton S, Sheridan EA, Ramsay M, Johnson AP. 2017. Resurgence of scarlet fever in England, 2014-16: a population-based surveillance study. *Lancet Infect Dis* 10.1016/S1473-3099(17)30693-X.

246. Chalker V, Jironkin A, Coelho J, Al-Shahib A, Platt S, Kapatai G, Daniel R, Dhami C, Laranjeira M, Chambers T, Guy R, Lamagni T, Harrison T, Chand M, Johnson AP, Underwood A, Scarlet Fever Incident Management Team. 2017. Genome analysis following a national increase in scarlet fever in England 2014. *BMC Genomics* **18:**224 http://dx.doi.org/10.1186/s12864-017-3603-z.

247. Turner CE, Pyzio M, Song B, Lamagni T, Meltzer M, Chow JY, Efstratiou A, Curtis S, Sriskandan S. 2016. Scarlet fever upsurge in England and molecular-genetic analysis in North-West London, 2014. *Emerg Infect Dis* **22:**1075–1078 http://dx.doi.org/10.3201/eid2206.151726.

248. Grundmann H, Hori S, Tanner G. 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J Clin Microbiol* **39:**4190–4192 http://dx.doi.org/10.1128/JCM.39.11.4190-4192.2001.

249. Linke C, Young PG, Kang HJ, Bunker RD, Middleditch MJ, Caradoc-Davies TT, Proft T, Baker EN. 2010. Crystal structure of the minor pilin FctB reveals determinants of group A streptococcal pilus anchoring. *J Biol Chem* **285:**20381–20389 http://dx.doi.org/10.1074/jbc.M109.089680.

250. Facklam R, Beall B, Efstratiou A, Fischetti V, Johnson D, Kaplan E, Kriz P, Lovgren M, Martin D, Schwartz B, Totolian A, Bessen D, Hollingshead S, Rubin F, Scott J, Tyrrell G. 1999. *emm* typing and validation of provisional M types for group A streptococci. *Emerg Infect Dis* **5:**247–253 http://dx.doi.org/10.3201/eid0502.990209.